

SEARCH REQUEST FORM

Scientific and Technical Information Center

Access DB# _____

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 301 _____ Serial Number: _____
 Mail Box and Bldg./Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

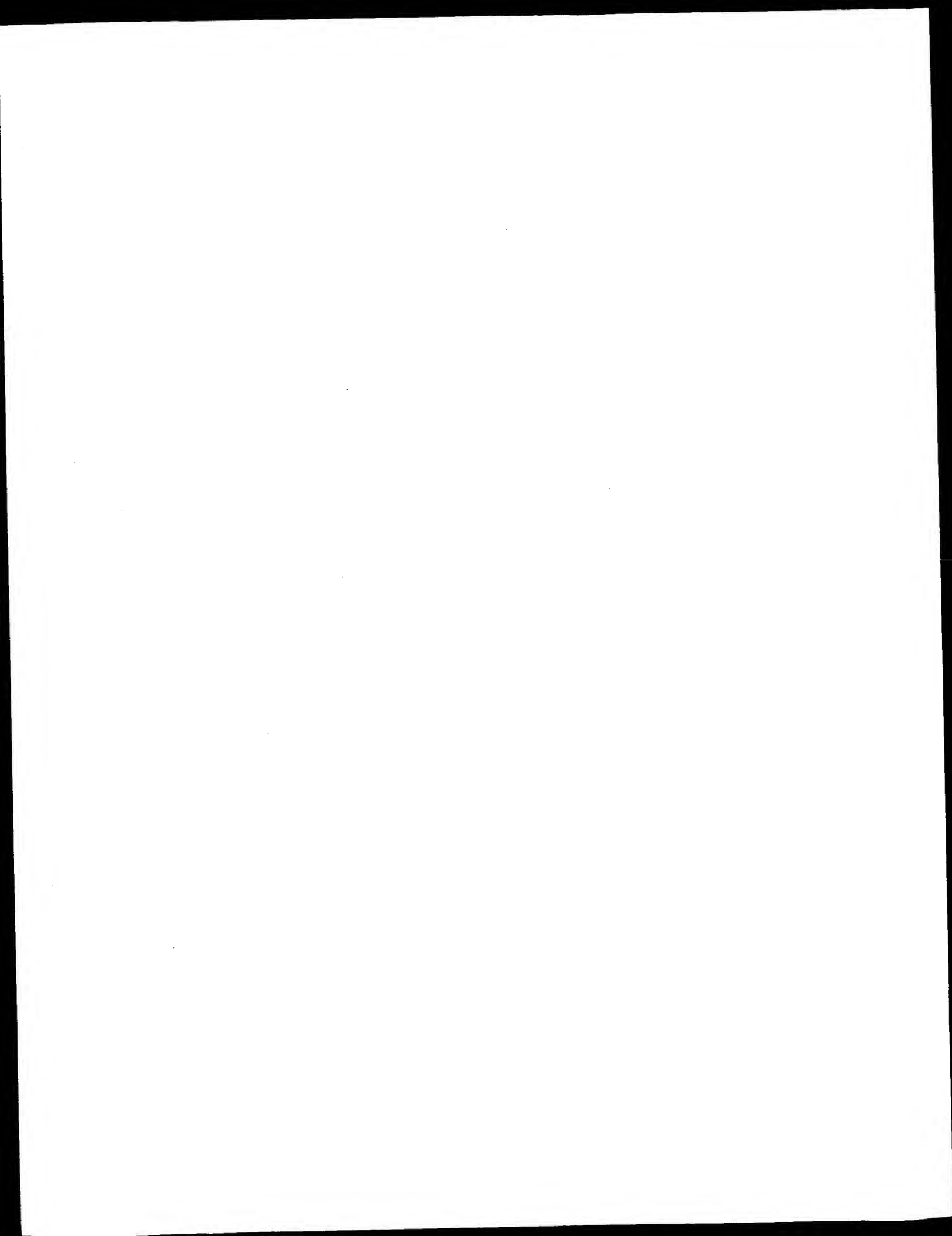
**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>Toby Port</u>	NA Sequence (#) <u>13</u>	STN _____
Searcher Phone #: <u>308-3534</u>	AA Sequence (#) _____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel Orbit _____
Date Searcher Picked Up: <u>10/24</u>	Bibliographic _____	On Line _____
Date Completed: <u>10/25</u>	Litigation _____	Lexis Nexis _____
Searcher Prep & Review Time: <u>15</u>	Fulltext _____	Sequence Systems <u>CS</u>
Client Prep Time: _____	Patent Family _____	WWW Internet _____
Online Time: <u>15</u>	Other _____	Other (specify) _____

PT 10/25/01

BEST AVAILABLE COPY



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 09:56:55 ; Search time 5479.82 Seconds
(without alignments)
2227.014 Million cell updates/sec

Title: US-09-462-955-1
Perfect score: 1291

Sequence: 1 cgcacaaacctctgctaag.....ggacggctgagtgatctgg 1291

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_estl1:*
2: gb_estl2:*
3: gb_estl3:*
4: gb_estl4:*
5: gb_estl5:*
6: gb_estl6:*
7: gb_estl7:*
8: gb_estl8:*
9: gb_estl9:*
10: gb_estl10:*
11: gb_estl11:*
12: gb_estl12:*
13: gb_estl13:*
14: gb_estl14:*
15: gb_estl15:*
16: gb_estl16:*
17: gb_estl17:*
18: gb_estl18:*
19: gb_estl19:*
20: gb_estl20:*
21: gb_estl21:*
22: gb_estl22:*
23: gb_estl23:*
24: gb_estl24:*
25: gb_estl25:*
26: gb_estl26:*
27: gb_estl27:*
28: gb_estl28:*
29: gb_estl29:*
30: gb_estl30:*
31: gb_estl31:*
32: gb_estl32:*
33: gb_estl33:*
34: gb_estl34:*
35: gb_estl35:*
36: gb_estl36:*
37: gb_estl37:*
38: gb_estl38:*
39: gb_estl39:*
40: gb_estl40:*
41: gb_estl41:*
42: gb_estl42:*
43: gb_estl43:*
44: gb_estl44:*
45: gb_estl45:*
46: gb_estl46:*
47: gb_estl47:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	40	3.1	512	224	AQ121628 HS_3089_A
2	38.2	3.0	413	3	AA171666 zo94g01.r
3	38.2	3.0	516	8	AA516896 vhr7e09.r
4	38	2.9	187	133	BB423647 BB423647
5	38	2.9	698	141	BE882337 BE882337
6	38	2.9	2275	141	AF034173 AF034173
7	36.8	2.9	498	231	AQ615572 HS_5144_B
8	36.8	2.9	518	121	AW818171 CM1-ST027
9	36.4	2.8	431	102	AT816449 AT816449
C 10	36.4	2.8	459	103	AT79137 AT79137
C 11	36.4	2.8	538	168	BF724252 BF724252
C 12	36.4	2.8	914	154	BG499317 BG499317
C 13	36	2.8	507	244	AZ444169 AZ444169
C 14	35.8	2.8	678	145	BF152590 BF152590
C 15	35.6	2.8	318	224	AQ089903 HS_3001_A
C 16	35.6	2.8	1101	219	CNS016H0 CNS016H0
C 17	35.4	2.7	460	104	AI942816 AI942816
18	35.4	2.7	488	219	CNS0000X CNS0000X
19	35.4	2.7	955	137	BE603735 BE603735
C 20	35.2	2.7	345	243	AZ401897 AZ401897
21	35.2	2.7	793	30	AV400883 AV400883
22	35.2	2.7	848	233	AQ746046 AQ746046
23	35	2.7	644	31	AV593739 AV593739
24	34.6	2.7	442	251	AZ911207 AZ911207
25	34.6	2.7	487	228	AQ418525 AQ418525
26	34.6	2.7	555	228	AQ417598 AQ417598
C 27	34.4	2.7	594	243	AZ397924 AZ397924
C 28	34.4	2.7	788	174	BG114850 BG114850
C 29	34.2	2.6	477	136	BE498734 BE498734
C 30	34.2	2.6	779	152	BG320945 BG320945
C 31	34.2	2.6	842	137	BE558794 BE558794
C 32	34.2	2.6	878	151	BF626764 BF626764
C 33	34.2	2.6	934	164	BF214212 BF214212
C 34	34	2.6	442	225	AQ228962 AQ228962
C 35	34	2.6	443	223	AQ010299 AQ010299
C 36	34	2.6	1101	219	CNS000DBS CNS000DBS
C 37	33.8	2.6	172	133	BB423076 BB423076
C 38	33.8	2.6	283	132	BB353778 BB353778
C 39	33.8	2.6	291	135	BB485914 BB485914
C 40	33.8	2.6	410	233	AQ07176 AQ07176
C 41	33.8	2.6	604	219	AG019820 AG019820
C 42	33.8	2.6	749	241	AZ331069 AZ331069
C 43	33.8	2.6	772	229	AQ488982 AQ488982
44	33.8	2.6	894	221	CNS032CO CNS032CO
C 45	33.8	2.6	1174	140	BE779656 BE779656

ALIGNMENTS

RESULT	1	Query Match	Length	DB ID	Description
AQ121628/c					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					

JOURNAL MEDLINE COMMENT

scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
Plate: 3089 row: C column: 5
Class: BAC ends
High quality sequence stop: 512.

Location/Qualifiers
1. .512

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3089 Col=5 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 174 a 85 c 115 g 136 t 2 others
ORIGIN

FEATURES

source

Query Match

Best Local Similarity

Matches

Conservative

Score

DB

Length

512

Indels

Gaps

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

JOURNAL MEDLINE COMMENT

scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
Plate: 3089 row: C column: 5
Class: BAC ends
High quality sequence stop: 512.

Location/Qualifiers
1. .512

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3089 Col=5 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 174 a 85 c 115 g 136 t 2 others
ORIGIN

FEATURES

source

Query Match

Best Local Similarity

Matches

Conservative

Score

DB

Length

512

Indels

Gaps

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 755 agaccacaaacgaatttaattcctcgaatgaccaggtgttaatttagaagatttaatt 814
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 168 AGGATTAAACCAATTTGATCTCATGTGACACCAAGCAAAATATAGTAAAGATT 227
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 815 atgacctgttagatgtgttaagaacaggcgattca 850
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 228 CTGCTCAGAGTAAGTCTTAAACACCTGCCCAATCA 263

RESULT 9
 A1816449/c
 LOCUS A1816449 431 bp mRNA EST 09-JUL-1999
 DEFINITION au46f05.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
 IMAGE:2517825 5' similar to TR:Q19737 Q19737 F2E10.5 ; mRNA
 sequence.
 ACCESSION A1816449
 VERSION A1816449.1 GI:5431995
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 431)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,
 J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 TITLE WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 407.

FEATURES
 source
 1..431
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2517825"
 /clone_lib="Schneider fetal brain 00004"
 /sex="male"
 /tissue_type="frontal lobe"
 /dev_stage="5 months post-conception"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);
 Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
 prepared from human fetal brain tissue. 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence:
 5'-GAGAGAGAGAGAGCTCAAGGATCTTTAAATTAATTAATCCCCCCCCCCC-3'
 and 3' adaptor sequence:
 5'-GAGAGAGAGAGAGCTCGAGTTTCTTTTCTTTT-3'. The library was
 size-selected for >0.5 kb inserts and has an average
 insert size estimated at 1.2 kb. This library was
 constructed using the CAP-trapper method for full-length
 enrichment and has not undergone amplification. Library
 was constructed by Dr. Claudio Schneider (LNCIB-Area
 Science Park, Trieste, Italy)."
 BASE COUNT 44 a 172 c 159 g 56 t
 ORIGIN

Query Match 2.8%; Score 36.4; DB 102; Length 431;
 Best Local Similarity 59.8%; Pred. No. 3.2;
 Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 40 agccgcgggtagtattaccgcgggtctcccaacctctgtctaaccccgcttgcta 99
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 346 AGCAGCCAGGTCCAGTAGCTGCAGCGCGGCTCGAGCAGCAGCGCCGCCGCGCTG 287
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 100 taatggtgttcattccattccgctgtgtctcactctgaac 141
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 286 TAGCGGTGCTCTCTCAGTCGCCGAGCTGCGCGCGCTCAGC 245

RESULT 10
 A1879137/c
 LOCUS A1879137 459 bp mRNA EST 23-AUG-1999
 DEFINITION au55a12.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone
 IMAGE:2518654 5' similar to TR:Q19737 Q19737 F2E10.5 ; mRNA
 sequence.
 ACCESSION A1879137
 VERSION A1879137.1 GI:5553186
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 459)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,
 J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 TITLE WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 438.

FEATURES
 source
 1..459
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2518654"
 /clone_lib="Schneider fetal brain 00004"
 /sex="male"
 /tissue_type="frontal lobe"
 /dev_stage="5 months post-conception"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript SK (Stratagene);
 Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
 prepared from human fetal brain tissue. 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence:
 5'-CAGAGAGAGAGAGCTCAAGGATCTTTAAATTAATTAATCCCCCCCCCCC-3'
 and 3' adaptor sequence:
 5'-CAGAGAGAGAGAGCTCGAGTTTCTTTTCTTTT-3'. The library was
 size-selected for >0.5 kb inserts and has an average
 insert size estimated at 1.2 kb. This library was
 constructed using the CAP-trapper method for full-length
 enrichment and has not undergone amplification. Library
 was constructed by Dr. Claudio Schneider (LNCIB-Area
 Science Park, Trieste, Italy)."
 BASE COUNT 78 a 146 c 138 g 97 t
 ORIGIN

Query Match 2.8%; Score 36.4; DB 103; Length 459;
 Best Local Similarity 59.8%; Pred. No. 3.2;
 Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 40 agccgcgggtagtattaccgcgggtctcccaacctctgtctaaccccgcttgcta 99
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

DB 172 AGCAGCCAGGTCCTCAGTAGAGCTGTCAAGCGCGGCTCGACGACGAGACGCCGGCGGCTG 113

Qy 100 taaatgggttcctccattgcgcgctgggtgcttcaactotgaac 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 112 TAGCGGTGCTCTCCCTCCAGTCCGCGCAGCTCGCGCGGCTCAGC 71

RESULT 11
BF724252/c

LOCUS BF724252.1 538 bp mRNA EST 05-JAN-2001

DEFINITION bx02f09.y1 Human Iris cDNA (Un-normalized, unamplified): BX Homo sapiens cDNA clone bx02f09 5', mRNA sequence.

ACCESSION BF724252

VERSION BF724252.1 GI:12040161

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 538)

AUTHORS Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.

TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics

JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 02 . Row: f column: 09
Seq primer: M13Rpl reverse primer (ABI).

FEATURES
source Location/Qualifiers
1..538
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clones="bx02f09"
 (clone_lib)="Human Iris cDNA (Un-normalized, unamplified):
 BX"
 /tissue_type="Iris"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /note="Organ: Eye; Vector: pCMVSPORT6; Post-mortem iris tissue was pooled from 10 individuals ranging in age from 4-80 years and RNA was extracted. From this pooled sample an aliquot of 60ug of total RNA yielded 2.17ug of mRNA. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript plasmid System full details of which are contained in the manufacturer's instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor [5'-pgACTAGTTCTAGATCGGCGGCCGC(T)15-3']. Not I blunt end inserts were cloned into the Not I/EcoRI sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT 79 a 200 c 175 g 84 t

ORIGIN

Query Match 2.8%; Score 36.4; DB 168; Length 538;
Best Local Similarity 59.8%; Pred. No. 3.4;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 40 agcgcgggggcttagttaccgcccgggctcccccaacctctgtctaaccgcgttgagcta 99
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 331 AGCAGCCAGGTCCTCAGTAGAGCTGTCAAGCGCGGCTCGACGACGAGACGCCGGCGGCTG 272

Qy 100 taaatgggttcctccattgcgcgctgggtgcttcaactotgaac 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 271 TAGCGGTGCTCTCCCTCCAGTCCGCGCAGCTCGCGCGGCTCAGC 230

Query Match	2.8%;	Score 36;	DB 244;	Length 507;
Best Local Similarity	58.3%;	Pred. No. 4.5;		
Matches	63;	Conservative	0;	Mismatches 45; Indels 0; Gaps 0;
Y 894	tgtgcattgtaactgcgtatttggccaatgtcctgcgtgattatttgaataatcagcaggacag	953		
b 364	TATTCATGAAC*AGAATTTGGGAAGCATTTCTCTATTATTAATGATCAGTAGCCATAT	305		
Y 954	aataaaactgtggaataattttaaagtatgtgctcatctaaatacaccaa	1001		
b 304	GATATTGTTGTC*TGAGGTGGAATCATTTTATTTTAAATAACAACA	257		
RESULT 14				
FI52590/c				
OCUS	BF152590	678 bp	mRNA	EST 29-DEC-2000

[illegible]

	Query Match	2.8%	Score 35.6;	DB 224;	Length 318;
	Best Local Similarity	54.6%;	Pred. No. 5.2;		
	Matches 71;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;
QY	832	gttaagaacagggcattcagtcgcgacaaatacgaacccttagtattctctgggttcgac	891		
DBb	212	GTTCGTGGCCAGGGCANTCAGCGAGGACAGGAATAAAGGTATTATTATTAGACAGAG	153		
QY	892	catgtgcagtgtactcgtatttggccaatgtcctgcctgatttggaaatacagcagggac	951		
DBb	152	GATTTCAAAATTGCCCTGTTTGCAGATGACATCACTGTATATCTAGAAAAACCGCATTTGTC	93		
QY	952	agaataaaac	961		
DBb	92	TCAGTCCCAAC	83		

Search completed: October 24, 2001, 13:13:35
Job time: 11800 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:19 ; Search time 180.6 Seconds
(without alignments)
818.671 Million cell updates/sec

Title: US-09-462-955-1_COPY_211_991
Perfect score: 781
Sequence: 1 gttggggacgaggtgcacc.....ttaaagtatgtgtcatctaa 781

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents.NA.*
- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
 - 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
 - 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
 - 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
 - 5: /cgn2_6/ptodata/2/ina/PCtUS_COMB.seq.*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	162.6	20.8	1096	2	US-08-418-071-4
2	142.6	18.3	1091	2	US-08-418-071-5
3	137.8	17.6	1106	2	US-08-418-071-3
4	108	13.8	1022	4	US-08-793-634B-2
5	99.2	12.7	1017	4	US-08-793-634B-6
6	62.8	8.0	1110	1	US-08-202-186-14
7	59.6	7.6	982	3	US-08-973-068-28
8	59.6	7.6	1110	1	US-08-202-186-11
9	59.6	7.6	1111	1	US-08-202-186-9
10	59.6	7.6	1111	1	US-08-202-186-12
11	59.6	7.6	1111	1	US-08-202-186-24
12	58.6	7.5	1103	1	US-08-202-186-16
13	58.6	7.5	1103	1	US-08-202-186-18
14	58	7.4	1109	1	US-08-202-186-13
15	57	7.4	1111	1	US-08-202-186-15
16	57	7.3	1104	1	US-08-202-186-17
17	56.4	7.2	1111	1	US-08-202-186-10
18	46.4	5.9	287	2	US-08-418-071-1
19	46.4	5.9	300	2	US-08-418-071-2
20	37.2	4.8	7218	1	US-08-232-463-14
21	31.8	4.1	498	1	US-07-781-254A-14
22	31.8	4.1	1338	1	US-07-781-254A-6
23	31.8	4.1	1338	1	US-07-781-254A-7
24	31.8	4.1	1863	1	US-07-781-254A-4
25	31.8	4.1	1863	1	US-07-781-254A-5
26	31	4.0	1212	3	US-09-046-578-5
27	30.4	3.9	1212	3	US-09-046-578-1

28	30	3.8	1437	3	US-08-724-814-15	Sequence 15, Appli
C 29	29.8	3.8	951	1	US-08-671-525B-1	Sequence 1, Appli
C 30	29.8	3.8	951	1	US-08-672-109B-1	Sequence 1, Appli
C 31	29.8	3.8	951	1	US-08-842-045-1	Sequence 1, Appli
C 32	29.8	3.8	951	1	US-08-842-238-1	Sequence 1, Appli
C 33	29.8	3.8	951	1	US-08-629-335B-1	Sequence 1, Appli
C 34	29.8	3.8	1633	1	US-07-866-979-5	Sequence 5, Appli
C 35	29.8	3.8	1633	2	US-08-466-906B-5	Sequence 5, Appli
C 36	29.8	3.8	1633	3	US-08-706-281A-5	Sequence 5, Appli
C 37	29.8	3.8	1633	4	US-09-201-746-5	Sequence 5, Appli
C 38	29.4	3.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
C 39	29.4	3.8	8920	4	US-09-150-741-1	Sequence 1, Appli
C 40	29.2	3.7	1209	6	5352575-4	Patent No. 5352575
C 41	29.2	3.7	1596	5	PCT-US94-11328A-3	Sequence 3, Appli
C 42	28.8	3.7	10607	1	US-08-078-090-3	Sequence 3, Appli
C 43	28.6	3.7	1593	2	US-08-524-828-2	Sequence 2, Appli
C 44	28.6	3.7	1593	2	US-08-975-114A-2	Sequence 2, Appli
C 45	28.6	3.7	1593	3	US-08-849-281A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BTV)
US-08-418-071-4

[illegible]

Qy	481	ctcaagccgcagctggtctcacacatggtgggaaccagaaaagcagctattgtaccagtaac	540
Db	636	TTAAAAACCTGGATGGGATATATCAACGGTGGAAAGACGTCGGATATCATGCACATCAT	695
Qy	541	atcgaggagcccaaacgaaatttaactcctcgatgtaccacagtgtaatttagagatttta	600
Db	696	ACGATGGATCTGTATATCATCTGGATTATGTGATATCCCCCAGAGTCAPTCCAGATTATCTG	755
Qy	601	aattatgcccgtgttagaatdgtttaagaacaggccattcagttcggacaaatacgaac-	659
Db	756	AAPTATGGCGGTATAGAACCAAAATTAAGAAATAGAGTTTTTAATAAATACAAAATACGAACCA	815
Qy	660	--cctagtattctctggttccgaccatgfgcatgtactcgtatttgccaatgtcctgct	717
Db	816	TGTGTGATTACAAAAGATGCACAAATGTCCATGTAAATGTTATGGCAAAATGTGTGCT	875
Qy	718	gattattgaaatacagcaggagcagaataaaactgtggaattttaaagtatgtgtcat	777
Db	876	GATTATTGAAATTTTCAGAAGATAGAAATAAAAATAAATTAATTGTTTCAGAAAAGAAACTT	935
Qy	778	c	778
Db	936	C	936

```

RESULT      3
US-08-418-071-3
; Sequence 3, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

```

APPLICANT: Marchall, Jerry S.
TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793.634B
FILING DATE: June 9, 1997

ATTORNEY/AGENT INFORMATION:
NAME: DiGiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
8-793-6348-2

Query Match	13.8%;	Score 108;	DB 4;	Length 1022;
Best Local Similarity	49.3%;	Pred. NO.	3e-26;	
Matches 278.	Conservative	0.	Mismatches	370;
			Indels	18;
			Gaps	3;

1	QY	gttgggacgaggttgcaacttcactcggcacaacgcaacctccaaggattcaccactg	60
178	Db	gtttgtcgcgcaaaaactgcaactacttcgcagaaaacacctccaggattgtatcggttc	237
61	QY	aagaccggtcggcgactgcaaggattgaagactgtctctgggaatgacagattccactg	120
238	Db	agacacaaaattcgctgtggtgaatgaagaaaatttggtaatc--gagctcactcgg	294
121	QY	gagccgaccogtggttccgacgaacagaatagagactactcttcgaaggaaacgggtgctt	180
295	Db	gaatttcgcgagcagcgacttcagaaatcgcgattattcgtctaaagaaacccctaatt	354
181	QY	ctcaagcacagagtcgccgactcgtcctcgagtcgaaggccacgattggcccaacgattt	240
355	Db	tcctgaatttgggattcccgctcatgaagggttcacacagcggacgatggagatttat	414
241	QY	gctgaggacactgatgaactccgcctgggaagccacgagcggaatccgaagaatcgtgtga	300
415	Db	gaagaggaatcccgaaagaaatgcgaattgaagagatccagatactgctcttcgatgaagcg	474
301	QY	caaggactcgtgggaatggacaagatggccgctgaaaatcggcttcccatctccatat	360
475	Db	agaaaattgaagagggaattattgttctgttatgattttcgaanaactccgtcca-----	528
361	QY	cacaattggcagctgaagtgcgtcgtcgatcgagagccagcgagcatcgacaatc	420
529	Db	-----tggcacaattgagcttcacgaggaatttaattggcggaaccagatgatcgcgactatc	582
421	QY	ctctggatatcgacgagacgagacgagacggaagtcggtgttgcacaaatctctcgga	480
583	Db	atcttgggtctatggtttcagacggagagagaagaacacgagcttcgcgaaggaatt---a	639
481	QY	ctcaagcccgactggttctacatctggtggaaacgaagaagcgattgtaccaggtac	540

[illegible]

RESULT 5
US-08-793-634B-6
; Sequence 6, Application us/08793634B
; Patent No. 6211431
; GENERAL INFORMATION:
; APPLICANT: Boevink, Petra C.
; APPLICANT: Surin, Brian P.
; APPLICANT: Keese, Paul K.
; APPLICANT: Chu, Paul W.G.
; APPLICANT: Waterhouse, Peter M.
; APPLICANT: Khan, Rafiqul I.
; APPLICANT: Larkin, Philip J.
; APPLICANT: Taylor, William C.
; APPLICANT: Marshall, Jerry S.
; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES

NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,634B
FILING DATE: June 9, 1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10530
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1017 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-793-634B-6

Query Match	12.7%	Score 99.2;	DB 4;	Length 1017;
Best Local Similarity	54.1%	Pred. No. 2.5e-23;		
Matches 225; Conservative		0; Mismatches 188;	Indels 3;	Gaps 1;

Qy	361	cacaattgcagcttgaagtgtctctgcgatcgagagccagcgacgatcgacaac	420
Db	498	CGCCCCCTGGCAAGAAGAAGTCATCTTCTAGAGGAACACGATATAGAACGATA	557
Qy	421	ctctggatatcgacagagcagagagacggaagtcggttgcccaaatctatctcgga	480
Db	558	ATCTGGGHTGATGACCTGCTGTAATGAAGGCAAACTACATTTGCAAGACATCTGTCA	617
Qy	481	ctcaagcccgactggttctcacatgtgggtgaaccagaaagacattatgcaggatc	540
Db	618	TTGAAGATGTTGGGTTATCTCCCTGGAGGAACACACAGATATGATGATCTTGG	677
Qy	541	atcagagaccacaacaaathtaactctgatgtaccaggtgtgaatttagagtattta	600
Db	678	ACTGCTGAGCCTAAGATAAATGGGTATTGACATACCAGAGTTAGTTCAGAGTATGTG	737
Qy	601	aattatccctgttgaagtgttaagaacacagggcaatcagttcggagcaaatcacaaac	659
Db	738	AATTTATGCTGTATAGAACACAGGTTAGAATAGGGTAATGGTGAATACTAAGTATGAGCA	797
Qy	660	--cctagttatcttgggttcgaccatgtgcattactcgtatttgcaatgtctcgtcct	717
Db	798	TGCTAATGCGGGATGATAATCATCTCTGTTCACTAATGTCTGTTCCAAAGTACTCCCA	857
Qy	718	gattattgaaatcagcagggcagcaataaactgtggaaatatttaagtatgtg	773
Db	858	GATTTGGGAAATTAAGTGAAGATAGAAATAAAATTAATTCGTGTGTGAAAACCTCG	913

```

6
RESULT
US-08-202-186-14
; Sequence 14, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-14

```

Query Match 8.0%; Score 62.8; DB 1; Length 1110;

[illegible]

```

RESULT 7
US-08-973-068-28
; Sequence 28, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Daile, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; CURRENT FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; EARLIER FILING DATE: 1996-05-31
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 982
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-28

```

Query Match	7.6%	Score 59.6;	DB 3;	Length 982;
Best Local Similarity	51.3%;	Pred. No. 3.9e+10;		
Matches 193;	Conservative 0;	Mismatches 174;	Indels 9;	Gaps 2;
QY 378	agtgcgtgctcgatcgcagagccagcgagcatcgacacatcctctgatatgcgacg	437		
Db				
QY 53	agagtgaggagaatcatctggcgagccatgtcatcggagaaataatttggtctatggcc	112		
Db				
QY 438	agacggaggagacgggaagtcgctgtttgccaaatctctggactcaagcccgactggtt	497		
Db				
QY 113	aaatggaggagaagaaagacacacgctatgcgaaacatctaatgaagacgagaaatgcgtt	172		
Db				
QY 498	ctacacatgtggtggaaccagaaagagcgtatgtaccagtcacatcgagaccaccaaacg	557		
Db				
QY 173	ttattctcagaggaaaataatctgggataatgtagactgtaataatcgaggatatt---	229		
Db				
QY 558	aaatttaactctcgatgaccaggtgtgaattttagagtatttaaatattgcctgttaga	617		
Db				
QY 230	---tggtatatattgatactcgaatgcgaagagattttaaatattaggtattaga	286		
Db				

Qy	618	atgtgttaagaacaggcgatttcgacaaatacgaaccccttagttatcttgggtt	677
Db	287	ggaatttaagaatgaataataccaagcggaatatgaaccgctttgagatag---t	343
Qy	678	cgaccattgcgtactcgtatttgcgaatgcctgcctgattattgaaaatcagcag	737
Db	344	agaatgtgcgaagtcattgtaatggctaacttcctccgaaggaagaatctttctga	403
Qy	738	ggacagataaaactg	753
Db	404	agatcgataaagtg	419

Sequence 11, Application US/08202186
Patent No. 5756708
GENERAL INFORMATION:
APPLICANT: KARAN, Mirko
APPLICANT: BURNS, Thomas M.
APPLICANT: DALE, James L.
APPLICANT: HARDING, Robert M.
TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1110 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-11

	Matches	193; Conservative	0; Mismatches	174; Indels	9; Gaps
Qy	378	agtgctgtctgcatcgagagccagcgagcgcacatacctctgcatatcgcgacg	43		
Db	584	AGAGGTGGAGAAATATCATCGGCACGCCATGTCTACGGAGAAATAATTTGGGTCTATGGACC	64		
Qy	438	agcggagagacgggaagtcggtgttgccaaatattctcggaactcaagccgcgactggtt	49		
Db	644	AAATGGAGGAGAGGAAGACAAACCTATGCAAAACATCTAATGAAGACGAGAAATCGGTT	70		
Qy	498	ctacacatgtgttggaaccagaaaggacgttatgtacacgtacatcgagaccaccaaacg	55		
Db	704	TTATTCTCCAGGAGGAAAATCAATTGGATATATGTAGACTGTATAAATTACGAGGATAT---	76		

Db 704 TTATTTCTCAGGAGGAAATCATTTGGATATATGTAGACTGTATAATTTACAGGATAT--- 760
QY 558 aaatttaattctcgtatgacccaggtgtaatttagagatatttaattatccctgttaga 617
Db 761 ----TGTATATTTGATATTCCTCAAGATGCAAGAGGATATTTAAATTTATGGTTATTAGA 817
QY 618 atgtgttaagacacgggcatcagttcggacaataacagaccccttagttatcttgggtt 677
Db 818 GGAATTTTAAAGATGAATAATTCAAACGGGGAATATGAACCCGTTTGAAGATAG---T 874
QY 678 gacacatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 737
Db 875 AGAATATGTCGAAGTTCATTTGATGCTGCTAACTTCCTCCGAGGAAGGAATCTTTTCCTCA 934
QY 738 ggacagataaaactg 753
Db 935 AGATCGAATAAAGTTG 950

RESULT 10
US-08-202-186-12
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-12

Query Match 7.6%; Score 59.6; DB 1; Length 1111;
Best Local Similarity 52.6%; Pred. No. 4.2e-10;
Matches 181; Conservative 0; Mismatches 154; Indels 9; Gaps 2;
QY 410 atcgacacatctctggtatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 469
Db 616 ATCGGAGATATTTGGTCTATGGCCCAAAATGAGGAGGAAGAAACACACATATGCCAA 675
QY 470 aatatctcggactcaagcccgactggttctacacatggtggaaccagagacgtat 529

Db 676 AACAACTAATGAAGACGAGGAATCGTTTATTTCTCCAGGGGAAAAATCATTTGATATAT 735
QY 530 tgtaccagatcatcgagagaccccaaaacgaatttaattcctcgtatgtaccaggtgtaatt 589
Db 736 GTAGACTGTATAATTTACAGGATAT-----TGTATATTTTGTATATTTCCAAAGATGCAAG 789
QY 590 tagagattttaaatatgacctgttagaatgtttaagaacagggcattcagttcggaca 649
Db 790 AGGATTTTAAATATTCGGTTATTAGAAGAAATTTAAGAAATGAATAATTTCAAAGCGGGA 849
QY 650 aatcagaccccttagttatcttgggttcacacatgcatgcatgcatgcatgcatgcatg 709
Db 850 AATATGAACCGGTTTGAAGATAG---TAGAATATATGCGAAGTCATTTGTAATGGCTAACT 906
QY 710 tcctgctgattattgaaaaatcagcagggacagataaaactg 753
Db 907 TCCTTCCGAAAGGAATCTTTTCTCAAGATCGAATAAAGTTG 950

RESULT 11
US-08-202-186-24
; Sequence 24, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; NAME/KEY: CDS
; LOCATION: 129..989
; US-08-202-186-24

Query Match 7.6%; Score 59.6; DB 1; Length 1111;
Best Local Similarity 51.3%; Pred. No. 4.2e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;
QY 378 agtgcgtctcgtatcgagagcagcagcagcagcagcagcagcagcagcagcagcagcag 437
Db 611 AGAGGTGGAGAATAATCATTCGGCGAGCCATGTCTATCGGAGAATAATTTGGTCTATGCC 670

Qy	438	aqacgaggaacaggaaagtccgtgtttgccaaatatctcggaactcaagcccgaactggtt	497
Dd	671	AAATGGAGGACGAAGAACAACCGTATCAAAACATCTAATGAAGACGAGAAATCCGT	730
Qy	498	ctacacatgtgttgaaaccagaaggaacgatattgtaccagtaactcgaaggacccaacaacg	557
Dd	731	TATTCTCAGGAGAAAATCATTTGGATATATCTAGACTGTAATAATACGAGGATAT---	787
Qy	558	aaatttaatactcogatgtaccagggtgaatttagagtatattaattatccccgtttaa	617
Dd	788	---TGTTATATTGATATCCAAGATSCAAAGAGGATATTTAAATATGGTTATTAGA	844
Qy	618	atgtgttaagaacaggggcatctcagtttcgacaaaatacgaaccccttagtatctctgggtt	677
Dd	845	GGAATTTAGAAATGGAATAATTCMAAGCGGAAATATGAACCCGTTTTGAAGATAG--T	901
Qy	678	cgaacctgtgcattactcgtattttgccaatgtccctgcctgattatttgaataatcagcag	737
Dd	902	AGAATATGCGAAGTCATTGATGGGTAACTTCCTTCGAAGGAAGAAATCTTTTCTGA	961
Qy	738	ggacagaataaaactg	753
Dd	962	AGATCGAATAAAAGTTG	977

```

RESULT      12
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent NO. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: KARAN, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

Query Match          7.5%; Score 58.6; DB 1: Length 1103;
Best Local similarity 52.5%; Pred. No. 9.le-10;
Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;

ov    410 atccgcacatcctcgattatcacgagacagcggaggaggaagtcccgttttgcca 469

```

[illegible]

```

RESULT 13
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-18

```

Query Match 7.5%; Score 58.6; DB 1; Length 1105;
Best Local Similarity 52.5%; Pred No. 9, 1e-10;
Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;
QY 410 atcgacacaatcctctgatatcgacagacgaggagacgggaagtcgggttttgcca 469

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 09:56:55 ; Search time 180.6 Seconds
(without alignments)
1353.271 Million cell updates/sec

Title: US-09-462-955-1
Perfect score: 1291
Sequence: 1 cgcacaaacctctgctaag.....ggacggtgagtgatctgg 1291

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents,NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_5/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_5/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_5/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	179.2	13.9	1096	2	US-08-418-071-4
2	163.4	12.7	1106	2	US-08-418-071-3
3	159.6	12.4	1091	2	US-08-418-071-5
4	112.4	8.7	1022	4	US-08-793-634B-2
5	99.2	7.7	1017	4	US-08-793-634B-6
6	62.8	4.9	1110	1	US-08-202-186-14
7	59.6	4.6	982	3	US-08-973-068-28
8	59.6	4.6	1110	1	US-08-202-186-11
9	59.6	4.6	1111	1	US-08-202-186-9
10	59.6	4.6	1111	1	US-08-202-186-12
11	59.6	4.6	1111	1	US-08-202-186-24
12	58.6	4.5	1103	1	US-08-202-186-16
13	58.6	4.5	1105	1	US-08-202-186-18
14	58	4.5	1109	1	US-08-202-186-13
15	58	4.5	1111	1	US-08-202-186-15
16	57	4.4	1104	1	US-08-202-186-17
17	56.4	4.4	1111	1	US-08-202-186-10
18	46.4	3.6	287	2	US-08-418-071-1
19	46.4	3.6	300	2	US-08-418-071-2
20	37.2	2.9	7218	1	US-08-232-463-14
21	31.8	2.5	498	1	US-07-781-254A-14
22	31.8	2.5	1338	1	US-07-781-254A-6
23	31.8	2.5	1338	1	US-07-781-254A-7
24	31.8	2.5	1863	1	US-07-781-254A-4
25	31.8	2.5	1863	1	US-07-781-254A-5
26	31.2	2.4	1768	4	US-09-347-594-3
27	31.2	2.4	1768	4	US-09-347-594-4

28	31	2.4	1212	3	US-09-046-578-5	Sequence 5, Appli
29	30.8	2.4	265	3	US-08-289-222E-5	Sequence 5, Appli
30	30.8	2.4	265	4	US-09-054-526B-5	Sequence 5, Appli
31	30.8	2.4	1767	4	US-09-347-594-1	Sequence 1, Appli
32	30.8	2.4	1767	4	US-09-347-594-2	Sequence 2, Appli
33	30.8	2.4	2272	1	US-08-482-577B-1	Sequence 1, Appli
34	30.8	2.4	2272	3	US-08-289-222E-2	Sequence 2, Appli
35	30.8	2.4	2272	4	US-09-218-176-1	Sequence 1, Appli
36	30.8	2.4	2272	4	US-09-054-526B-2	Sequence 2, Appli
37	30.4	2.4	1212	3	US-09-046-578-1	Sequence 1, Appli
38	30	2.3	1437	3	US-08-724-814-15	Sequence 15, Appl
39	30	2.3	3182	1	US-08-188-582-1	Sequence 12, Appl
40	30	2.3	3182	1	US-08-646-715-12	Sequence 12, Appl
41	29.8	2.3	951	1	US-08-671-525B-1	Sequence 1, Appli
42	29.8	2.3	951	1	US-08-672-109B-1	Sequence 1, Appli
43	29.8	2.3	951	1	US-08-842-045-1	Sequence 1, Appli
44	29.8	2.3	951	2	US-08-842-238-1	Sequence 1, Appli
45	29.8	2.3	951	3	US-08-629-335B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418.071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-4

Thu Oct 25 13:07:51 2001

```

Query Match      13.9%; Score 179.2; DB 2; Length 1096;
Best Local Similarity 52.4%; Pred. No. 2e-50;
Matches 499; Conservative 0; Mismatches 433; Indels 20; Gaps 4;

QY 40 agccgccccgtagtattacccccggcggtcccccaactctgtacaccccgcttgcta 99
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 AGGAGGAGCGGCTAGTATTACCGCTCTCTCGCACCTTCCTCGCACCTGACGTC-- 64

QY 100 taaatgggttcctcattccgctgggtgcttcaactctgaactcagagacagagagag 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 ---ATCATTTATCTCTTTTAAATGGTGTCTCACTCTCAATTTATCTCTCGCAGCGGAG 121

QY 160 gcgcggaattgttaagcgcatcgatcccttaacttgggtctatgtatgtggggag 219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 CGAGAAGACTTCTCGCTCTTCTGAAGGAGGAGGATGTTCACTACTCTGTCTGCTGGCGAC 181

QY 220 gagggtgacacttcaactgagacagacacacacacacacacacacacacacacac 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 GAAGTCTGCTCCCGCCACCGCCAGAGACACCTCCAGGGATATCTATCCCTGAANAATCA 241

QY 280 cggcgaactcaaggattgaagactgttcttgggaatgacagagattcacttgagccgacc 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 ATTCTGCTCGCGGATTGAAAAGAAGTATGG---CTCCGCTGCTCACTGGGAGATTGCG 298

QY 340 cgtggttcgacagacagaaatgagactactgttcgaaggagacgggtctctcagcac 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 AAAGGAAGTCAACAGAGATCGCATACTGTTCGAAGGAACCCCTAGTCTTCTGAACCTG 358

QY 400 ggaagtcacactcgtcgtgagtcgaagggccacgacttggcccaacgatttctgctgagaa 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 359 GGTACTCCGGTGTCTGCTGTTGCGAGAGAGCGCAAGCTTCTCATAGATTCAGAGAGAGC 418

QY 460 cctgtagaactccgctggaagaccagcgcgatataccgaagatgctgtgacacgagct 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 CCTGAGGAATGAAGATGACGATCCATCCAAAGTATCGCAGATGCTTGGCAGTGGAAATCA 478

QY 520 tcggtggaatggaagatggcgccgctgaaatccggttccatttccatcacatctgg 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 ATTAAAGATGCCAGAAATTAATTCGGAATGGGTTCAGCACTA-----AAAGAATGG 529

QY 580 cagcttgaagtgtgtctgcgactcgagagcgccagcgagatcgacacacacacacacac 639
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 CAAATAAATAATTAATCAACACATCGAAGGTGTTCTCGATGATCGAAGATATCATATCTGGGTA 589

QY 640 tgcgagacagacgagagagagagagagagagagagagagagagagagagagagagag 699
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 TACGCTCCCAACAGGAGGCGGAGAAAGTCAACCTTCGCAAGATATCTATCATTTAAACCC 649

QY 700 gactggttctacacatgtgtggaacacagagagagagagagagagagagagagagagag 759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 GGATGGGATATATCAACGGTGGAAAGACGCTCGGATATGATGATGATGATGATGATGATG 709

QY 760 ccaaacagaaatttaactcctcgatgtagccagagtagtaatttagagatttaaatatgcc 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 710 CCTGATAATCATTTGGATTTATGATATCCCAAGAGTCATTTCAGATTTATCTGGAATTTAGGC 769

QY 820 ctgttaaatgtgtaagaacagagcattcagttcgacacaaatcacgaacc---ccttagt 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 770 GTTATAGAACAAATTAAGATAGATTTTAAATAATACAAATACGAACCATGTTGTTGATT 829

QY 877 tatcttggttcgaccatgtgcatgtactcgtatttgcgaatgtcctgcctgtatttg 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 AGAAAGATGACAAAATGTCATGTAATTTATGTTATGCAAAATGTTGCTGATTTATGT 889

QY 937 aaatcagcagggacagaaataaaactgtggaattttaaagtatgtgtcatc 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 AAAATTTTCAGAGATAGAAATAAAATAATTAATTTGTTGAGAAAGAACTTC 941

```

RESULT 2

US-08-418-071-3

; Sequence 3, Application US/08418071

; Patent No. 5846705

```

GENERAL INFORMATION:
APPLICANT: Wu, Rey-Yuh
APPLICANT: You, Li-Ru
APPLICANT: Soong, Tai-Seng
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,071
FILING DATE: 06-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: DCB-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1106 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: genomic DNA
DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-3

```

```

Query Match      12.7%; Score 163.4; DB 2; Length 1106;
Best Local Similarity 50.6%; Pred. No. 4.3e-45;
Matches 454; Conservative 0; Mismatches 431; Indels 12; Gaps 2;

QY 83 ctaaccccgcttggtataaataagggttctcctcattccgctgctgcttctcactctgaact 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42 CTACCTCTGAACGCTGGAGATGTCACCTCCCTCTCTTAAGTGGTGGTCTCACTCTGAATT 101

QY 143 acgagacagagaaagagcgcgcaatgttgaaggcgcatcgatcccttaacttggct 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102 ACTCTCCGCGCACAGAGAGAAACTTCTCTCTCTTCTGAAGGAGGAGGATGTTCACT 161

QY 203 atgtatagttgggacagaggttgacacttcaacttcaacttcaacttcaacttcaacttca 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 162 ACGCTGTCTGCGGACGAAAGTCTGCTCGGCGCACGCGGCGGCGGCGGCGGCGGCGGCGG 221

QY 263 tccactgaagaccggtcgcgactgcaaggttgaagacttcttgggaatgacagga 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 222 TATCCCTGAAAAGAGAAATCCCGCTCGGCGGATTTGAAGAGAGATATGG---TTCCCGCTG 278

QY 323 ttcacctggagccgaccccggttctccgacgaacacacacacacacacacacacacacac 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 CTCACCTGGGAGATGTCAGAGAGAACGAGAGAAATTCGAAGTACTGTTTCAAAAGAAA 338

QY 383 ggggtcttcgagcagcgagtcctcctgagtcctcctgagtcctcctcctcctcctcctcct 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 CCCTAATCTCGAAATTAGGTTTCTGTTTAAATGGTCTTAATAAAGAGAAATATCGG 398

```

QY 443 aacgatttgtaggaacacatgatgaactccgctgggaagaccacgagcgagataccgaagat 502
 Db 399 AGATGCTTCTGCTCTCTGATCGCATGAAATTTGACACCCCTGAGATATTTCAACAGAT 458
 QY 503 gcggttgtagcagcgagctcggtggaatggacagatggcgctgaaatcccttcccat 562
 Db 459 -----ATCAATCTCTGTAATAGTAAATTAATAAATTTCAAGGAGGAGTTCGTTTCATCCTT 509
 QY 563 ttccatatacaaatggcagcttgaagtgctgctgctgacatcgagagccagcgagcagat 622
 Db 510 GCTCGATAGACCATGCGAGATTTCAATGACGAGGCAATTTGACGAGAAACCCGATGATC 569
 QY 623 gcaaatcctctgtagatgpggacgagacgagagagagcggaagtcgctgtttgccaat 682
 Db 570 GAAGCATCATCTGGGCTCTATGCTCTTATGCTAATGAGGGTAAATCAACATATGCCAAT 629
 QY 683 atctcggaactaacccgactggtttacacatggtggtggaacagaaagagcgtattgt 742
 Db 630 CACTAATCAAGAGGATTTGTTCTACACACGAGGGTGGGAAGAGGAGATATCTTATTCT 689
 QY 743 accgtagacatgagaccccaacgaaatttaacatcctcgatgacccaggttcagttcgacaat 802
 Db 690 CCTATGAGGAGGATCTTCACAGCATATATATTTGATATTTCTCTGTTGTAATCAGG 749
 QY 803 agtattaaattatgcctgttgaatgttgaacacagggcattcagtttcggacaat 862
 Db 750 ATTATTTAAATATGATGTATAGAGCATTAAGAGGATTAAGAGGATAGGTTATAGAGACTATAAT 809
 QY 863 acgaaccccttagttatcttggttcgacatgctgacatgctgacatgctgacatgctgacat 922
 Db 810 ACAACCCATTAAGATAGTAAATAGTAAATAGTAAATAGTAAATAGTAAATAGTAAATAGTAA 869
 QY 923 tgcctgattttgaaatcagcagagacagagagagagagagagagagagagagagagagagag 979
 Db 870 TGCCTGACTTCTGTAATCTCCGAAGATCGAATAAATAATATATTTATGCTGAAGAA 926

RESULT 3

US-08-418-071-5
 ; Sequence 5, Application US/08418071
 ; Patent No. 5846705
 ; GENERAL INFORMATION:
 ; APPLICANT: Wu, Rey-Yuh
 ; APPLICANT: You, Li-Ru
 ; APPLICANT: Soong, Tai-Seng
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECT
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10020
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 5.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/418,071
 ; FILING DATE: 06-APR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haley Jr., James F.
 ; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: DCB-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 596-9000
 ; TELEFAX: (212) 596-9090

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1091 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: subgenomic DNA
 ; DESCRIPTION: /desc "BBTV DNA II (clone 2)"
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 ; US-08-418-071-5

Query Match 12.4%; Score 159.6; DB 2; Length 1091;
 Best Local Similarity 51.5%; Pred. No. 8.2e-44;
 Matches 477; Conservative 0; Mismatches 429; Indels 20; Gaps 4;
 QY 66 cggctcccaaacctctgtaaccccgcttggtataaattgggtctccctccattccgcgtg 125
 Db 28 CCGCTCTCTCTCGCTTTCCTCTCGCCCTCGACGTCATCATTTATGCTCTCTTTTAAATG 87
 QY 126 gtgcttcactctaaactacgagacagaggaagcgcggaatgttgtaagcgcatcga 185
 Db 88 GTGCTTCACCTCTCAATTTATTCCTCCGACGCGGAGGAGACCTTTCTCGCTCTTCTGAA 147
 QY 186 gtcccttaactgtgctatgctatagttgggacgaggttgacacctcaactcgaacgcaacg 245
 Db 148 GGAGGAAGAGCTTAAATTACGCTCTCTCGGACGAGAGTCTCGTCCGAGCACCGTTCGGAA 207
 QY 246 acacctcaagagattccatccacttggaacccggtcggaactgcaagattgaagactgt 305
 Db 208 GCACCTCCAGGGATATCTATCTCTGAGAAATCTATTAAAGCTTGTGGATGAAGAAG 267
 QY 306 tcttggaatgacagattacactagacgacccggtggttcgacgacgaagaatagaga 365
 Db 268 GTACTC---TTCGAAGGCTACTGGGAGGCGGAGGAACTGATGAACAGATTCGCAG 324
 QY 366 ctactgttcgaaggaacgggtgtctctcgacgacgagtcctcgactcctcgagtgca 425
 Db 325 ATACTGTTCGAAGAA-----ACCCCTTGAACCTGGGTACTCCGGTGTCTCTGGTTCGA 379
 QY 426 agggccacgattggcccaacgattgtgagaaacctgatgaacctccgctggagacc 485
 Db 380 GAAGCCCAAGCTTCTCGATAGATTGAGAGAGAGCCCTCGAGGAATTTGAAGATGGCATCC 439
 QY 486 aggcgatacccaagatgcgttgtagcagagcttcggtggaatgggaagatggcgccg 545
 Db 440 ATCCAAGTATCGCAGATGCTTGGCAGTGAATCAATTAAAGATGCCAGATTAATTCGA 499
 QY 546 tgaataatccgtccattccatatacaaatggcagcttggaagtgctgtgctgacg 605
 Db 500 ATGGGTTTCAGAACTA-----AAAGAATGCAAAATAAATAATTCAACACATCGA 550
 QY 606 agagccagcgagatcgacaaatcctctgatatgagagacgagacgagagagcgga 665
 Db 551 AGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGCTCTCGCGAGGCGGAAGAAA 610
 QY 666 gtccgtgttgcaaatatctcggactcaagcccgactggttctacacatggtgtggaac 725
 Db 611 GTCAACCTTCGCAAGATATCTATCATTTAAACCTGGATGGGATATATCAACGCTGGA 670
 QY 726 cagaagagcgtattgtaccagtcacatcgagagcccaacaaatattatcctcgatg 785
 Db 671 GACGTCGGATATGATGACATCATTAACGATGGATCTCTGATAATCATTTGATATTGATAT 730
 QY 786 accaggtgtaatttagagatttaaatatgcccctgttagaagtggttaagaacagggc 845
 Db 731 CCCAGAAAGTCAATTCAGATTATCTGAATTTAGCGCTTATAGAACAAATTAAGATAG 790
 QY 846 attcagttcgaaraaatacgaacc---ccttagttatcttggttcgacctgcatgt 902

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Review #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIVE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1110 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-14

Query Match 4.9%; Score 62.8; DB 1; Length 1110;
Best Local Similarity 53.2%; Pred. No. 4e-11;
Matches 183; Conservative 0; Mismatches 152; Indels 9; Gaps 2;

7

US-08-973-068-28
; Sequence 28, Application US/08973068
; Patent No. 6127604

GENERAL INFORMATION:

APPLICANT: Dale, James Langham

APPLICANT: Harding, Robert M.

APPLICANT: Dugdale, Benjamin

APPLICANT: Hafner, Gregory F.

APPLICANT: Becker

TITLE OF INVENTION: INTERGEN

; FILE REFERENCE: 09657/002001

```

; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-9

```

```

Query Match          4.6%; Score 59.6; DB 1; Length 1111;
Best Local Similarity 51.3%; Pred. No. 4.8e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 588 agtgcgtcgtcgatcgagagccagcgagcgatcgacacatcctctggtatgagcg 647
Db 584 AGAGTGGAGAAATATCATGGCGCAGCCATCTCATCGGAGAAATAATTTGGTCTCTATGGCCC 643
QY 648 agacggaggagacgggaagtcggtgttgcgaatctctcgactcaagcccgactggtt 707
Db 644 AAATGGAGGAGAGAGAAAGCAACGATGTCAAACATCTAATGAACACGAGAAATGCGTT 703
QY 708 ctacacatggtggaaccgaagagcgatgttaccagtcacatcgagagcccaaacg 767
Db 704 TTATTCGCCAGGAGGAAATCATTTGGATATATAGACTCTATAATTACGAGGATAT--- 760
QY 768 aaatttaactctcgatgacccaggtgtaatttagagatttaaaattatgccctgtaga 827
Db 761 ---TGTTATATTGATATTCAGATGCCAAGAGGAGATTATTAATATATGGGTTATATAGA 817
QY 828 atgtgttaagacaggcgatcctcaggttcggaacaaatcagacccctctagtattcttgggtt 887
Db 818 GGAATTTAAGAAATGGAATAATTCAAAAGCGGAAATATGAACCCGTTTTCGAAGATAG---T 874
QY 888 cgaccatgtcagtcactcgtatttcccaatgctcctgcctgattatttgaataatcagcag 947
Db 875 AGAATATGTCGAGTCACTGTAATGCTAACTTCCTTCCCAAGGAAGGAATCTTTCTGA 934
QY 948 ggcagagaataaaactg 963
Db 935 AGATCGAATAAAGTTG 950

```

```

RESULT 10
US-08-202-186-12
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202.186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-12

```

```

Query Match          4.6%; Score 59.6; DB 1; Length 1111;
Best Local Similarity 52.6%; Pred. No. 4.8e-10;
Matches 181; Conservative 0; Mismatches 154; Indels 9; Gaps 2;

QY 620 atcgacacatcctctggtatgagcgagcgagcgagcgagcgagcgagcgagcgagcgat 679
Db 616 ATCGGAGAAATAATTTGGTCTATGGCCCAATGCGGAGAGAGAAAGCAACGTATGCAA 675
QY 680 aatatctcggaciaccaagcccgactggtttctacacatggtgtggaaccagaagcgat 739
Db 676 AACAACTAATGAACACGAGAAATGCGTTTATTCCTCCAGGGGAAATCATTTGGATAT 735
QY 740 tgtaccagtcacatcgagagcccaaacgaatttaactctcgtatgtaccaggtgtaact 799
Db 736 GTAGACTGTATAATTACGAGGATAT-----TGTTATATTGTATATTCAGATGCAAG 789
QY 800 tagagattttaaaattatgcccctgttagaatgtgttaagacagggcgatcagttcgga 859
Db 790 AGATTATTTAAATTAATGCGTTTATAGAGAAATTTAAGAAATGGAATAATTCAGAGCGGA 849
QY 860 aatacgaaccccttagttatcttgggttcgaccatgctgcatctcgtatttggcgaatg 919
Db 850 AATATGAACCCGTTTTCACATAG---TAGAATATGTCGAGTCAATGTGTAATGGCTAACT 906
QY 920 tctgctgattatttgaataatcagcagggagacagaataaaactg 963
Db 907 TCCTCCGAGGAGGAAGTAATCTTTCTCGAAGATCGAATAAAGTTG 950

```

```

RESULT 11
US-08-202-186-24
; Sequence 24, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIKI
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1111 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
FEATURE:
NAME/KEY: CDS
LOCATION: 129..989
US-08-202-186-24

Query Match 4.6%; Score 59.6; DB 1; Length 1111;
Best Local Similarity 51.3%; Pred. No. 4.8e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 588 agtctgtctgcatcgagagccagcgagcgatcgacacatcctctgdatcgcgacg 647
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 611 AGAGTGGAGAAATCATCGCGCAGCCATGTCATCGGAGAAATAATTTGGTCTATGCCCC 670

QY 648 agacggaggagacgggaagtcctgtttgccaaatctctcgactcaagcccgactggtt 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 671 AAATGGAGGAGAAAGCAAGCAACGATGTCACAAACATCTAATGAAGACGAGAAATGCGTT 730

QY 708 ctacacatgtgtgtaaccgaaagacgttatgtaccagtcacatcgagcccaaacg 767
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 731 TTATTCACGAGAAACATTTGGATATATGACATGTATAATTACGAGATAT--- 787

QY 768 aaattaatctctgatgtaccgggtgtaatttagagttattaaatagccctgttaga 827
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 788 ---TGTATTATTGATATTCAGATGCCAAGAGGATATTTAAATATGGGTATTAGA 844

QY 828 atgtttaaagacagggcattcagttcgcacaaatcgaaaccccttagttatcttggtt 887
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 GGAATTTAAGAAATGAATAATTCAAACGGGAAATATGAACCCGTTTGAAGATAG---T 901

QY 888 cgaccattgcatgtactcgtatttgccaatgtctcgtcctgatatattgaaaaatcagcag 947
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 902 AGAATATGCGAAGTCATGTTAATGGCTAACTTCCTCCGAAGGAAGGAATCTTTTCTGA 961

QY 948 ggacagaataaaactg 963
   || ||||| |||
DB 952 AGATCGAATAAAGTTG 977

RESULT 12
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA

```


Search completed: October 24, 2001, 10:00:12
Job time: 197 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:16 : Search time 393.52 Seconds
(without alignments)
1246.166 Million cell updates/sec

Title: US-09-462-955-1-copy_211_991
Perfect score: 781
Sequence: 1 gttgggacgaggtgtgacc.....ctaaagtatgtgtcatctaa 781

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N_Geneseq_0601.*
1:	/SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	162.6	20.8	1096	20	AAV71833
2	162.6	20.8	1096	21	AAA38950
3	145.4	18.6	886	20	AAA34687
4	142.6	18.3	1091	20	AAV71834
5	142.6	18.3	1091	21	AAA38947
6	137.8	17.6	1106	20	AAV71832
7	136.2	17.4	1106	21	AAA38946
8	108	13.8	1022	17	AAT13161
9	99.2	12.7	1017	17	AAT13165
10	80.4	10.3	593	20	AAA34686
11	62.8	8.0	1110	19	AAV24089
					BBTV DNA II clone
					Banana bunchy top
					DNA sequence of BB
					BBTV DNA II clone
					Banana bunchy top
					BBTV DNA I clone
					Banana bunchy top
					SCSV segment 2. S
					SCSV segment 6. S
					DNA sequence of BB
					Banana bunchy top

12	59.6	7.6	982	18	AAV49405	Banana bunchy top
13	59.6	7.6	1110	19	AAV24086	Banana bunchy top
14	59.6	7.6	1111	19	AAV24077	Banana bunchy top
15	59.6	7.6	1111	19	AAV24084	Banana bunchy top
16	59.6	7.6	1111	19	AAV24087	Banana bunchy top
17	58.6	7.5	1103	19	AAV24091	Banana bunchy top
18	58.6	7.5	1105	19	AAV24093	Banana bunchy top
19	58	7.4	1109	19	AAV24088	Banana bunchy top
20	58	7.4	1111	19	AAV24090	Banana bunchy top
21	57	7.3	1104	19	AAV24092	Banana bunchy top
22	56.4	7.2	1111	19	AAV24085	Banana bunchy top
23	46.4	5.9	287	20	AAV71830	Subgenomic fragmen
24	46.4	5.9	287	21	AAA38948	Subgenomic fragmen
25	46.4	5.9	300	20	AAV71831	Banana bunchy top
26	46.4	5.9	300	21	AAA38949	Subgenomic fragmen
27	45.8	5.9	936	22	AAF58252	Banana bunchy top
28	45.8	5.9	936	22	AAF58254	Oligonucleotide D1
29	45.8	5.9	936	22	AAF58257	Oligonucleotide D1
30	45.8	5.9	936	22	AAF58259	Oligonucleotide D1
31	45.8	5.9	936	22	AAF58262	Oligonucleotide D1
32	45.8	5.9	936	22	AAF58265	Oligonucleotide D1
33	42.4	5.4	936	22	AAF58252	Oligonucleotide D1
34	42.4	5.4	936	22	AAF58254	Oligonucleotide D1
35	42.4	5.4	936	22	AAF58257	Oligonucleotide D1
36	42.4	5.4	936	22	AAF58259	Oligonucleotide D1
37	42.4	5.4	936	22	AAF58262	Oligonucleotide D1
38	42.4	5.4	938	22	AAF58255	Oligonucleotide D1
39	37	4.7	244	22	AAF58238	Oligonucleotide D1
40	34.2	4.4	244	22	5053	Oligonucleotide D1
41	32.6	4.2	5053	19	AAV22816	Nucleotide sequenc
42	32.6	4.2	5055	19	AAV17236	DNA from a region
43	32.4	4.1	1149	21	AAA51838	Human melanocortin
44	32.4	4.1	1340	21	AAA51837	Human melanocortin
45	31.8	4.1	1863	12	AAQ14271	D-3 dopaminergic r

ALIGNMENTS

RESULT	1
AAV71833	
ID	AAV71833 standard; DNA; 1096 BP.
XX	AAV71833;
XX	10-FEB-1999 (first entry)
XX	BBTV DNA II clone (2-17) nucleotide sequence.
XX	Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
XX	Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
XX	Banana bunchy top virus.
XX	Key Location/Qualifiers
FT	TATA_signal 1..7
FT	stem_loop /*tag= a
FT	/*tag= b
FT	CDS 70..928
FT	/*tag= c
FT	/product= "ORF-V2 product"
FT	/transl_except= (pos:215..217, aa:Gly)
FT	533..538
FT	polyA_signal /*tag= d
FT	polyA_signal 799..804
FT	polyA_signal /*tag= e
FT	polyA_signal 907..912
FT	polyA_signal /*tag= f
FT	polyA_signal 1030..1035
FT	/*tag= ge
XX	US5846705-A.
PN	

Db 230 aaaaagagaatccgctcgccgattgaagaagatg---ttccgctgctcactgg 286
 Qy 121 gagcgcacccgtggttcgcagcaacaagaatagadactactgttcgaaggaacgggtgctt 180
 Pt 287 gagatgcagagggaacgagagagaaatcgaagtagctgttcgaaggaacccctaat 346
 Ps 181 ctgcagcagcaggtccgactgctcgtgagtcgaagggccacacatggtcccaacgat 240
 Cc 347 ctgcgaattaggttccctgtgttaagtgttctaalaagggaaataatcgagatggt 406
 Cc 241 gctgagggaacccgtgataaactccgctcggaagacccagcgcgataccgaagatggtgta 300
 Cc 407 gctcgtctcctgctgacgataaaattgaacagcctgagatatttcacagat----- 458
 Qy 301 cagcagcgtcgtggaatggacagatggcgcgtgaaatccgttcccatctccat 360
 Db 459 -atcaatctgtgaataagtttaaaaaattcaaggagaggttcgttcacctgctcgat 517
 Qy 361 cacaattggcagctggaatgctgctgagtcgagagccagcgcgagatcgacaatc 420
 Db 518 agaccatggcagatccaattgacggaggcaattgacgaggaaaccgagatgacgaagc 577
 Qy 421 ctctgagatcgagcagacgagagagagagagagagagagagagagagagagagagag 480
 Db 578 atctgggtctggttccttatgttaagtgagggaacaaatcaacatacggaagtcac 637
 Qy 481 ctcaagcccgactggtctcacatggtggtggaacacgaagagagagagagagagagag 540
 Db 638 aagaagattggtctcacaccgggtgggaagagagagagagagagagagagagagagag 697
 Qy 541 atcgaggaccccaaaagaaatttaactcctcgtatgaccaggtgtaatttaagattat 600
 Db 698 gacgaaggatctgacaagatagattggtgataattcctcgttgaatcagagattat 757
 Qy 601 aattatgacctgttagaattgttgaagacagagggaattcagtcgagacaaatcacga 660
 Db 758 aattatgagtaagagcattaaagagagagagagagagagagagagagagagagagag 817
 Qy 661 cttagttatcttggttcgacatgctgagatgactcgtatgctatgccaatgctcgtg 720
 Db 818 ataaagatagttgaattagtaaaatcacatgtaactgtaactgctgagagagagagag 877
 Qy 721 tatttgaataatcagcag 769
 Db 878 tctgtaaaatctccgaagatcgaaataaaatcatttattgtctgaagaa 926

RESULT 7
 AAA38946
 ID AAA38946 standard; DNA: 1106 BP.
 XX
 AC AAA38946;
 XX
 DT 23-AUG-2000 (first entry)
 XX
 DE Banana bunchy top virus related nucleotide sequence #1.
 XX
 KW Banana bunchy top virus; BBTV; detection; ds.
 XX
 OS Banana bunchy top virus.
 XX
 PN TW360710-A.
 XX
 PD 11-JUN-1999.
 XX
 PF 30-JUN-1994; 94TW-0106105.
 XX
 PR 30-JUN-1994; 94TW-0106105.
 XX
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX
 PI Wu R, You L, Song T;

XX WPI; 2000-316145/27.
 DR Two circular single-stranded DNAs associated with banana bunchy top
 XX virus and detection of the virus
 PT Claim 1; Page 1; 7pp; Chinese.
 XX
 PS The present invention describes two circular single-stranded DNAs
 CC associated with banana bunchy top virus (BBTV) and the encoded proteins.
 CC Also described is a method of detecting the BBTV virus using the
 CC polymerase chain reaction (PCR). The present sequence represents a
 CC BBTV related nucleotide sequence from the present invention.
 XX Sequence 1106 BP; 335 A; 226 C; 256 G; 289 T; 0 other;
 SQ
 Query Match 17.4%; Score 136.2; DB 21; Length 1106;
 Best Local Similarity 50.6%; Pred. No. 6.7e-36;
 Matches 389; Conservative 0; Mismatches 368; Indels 12; Gaps 2;
 Qy 1 gttgggacagaggttgacacttcaactggccaagacacactccaagattcaccacttg 60
 Db 170 gtcggcagcaagtcgctcgccacccggaagacacccctccagggatctatccctg 229
 Qy 61 aagaccggtcgccgactgcaaggattgaagactgttcttgggaatgacaggttaccctg 120
 Db 230 aaaaagagaatccgctcggtggtgaagagagagagagagagagagagagagagag 286
 Qy 121 gagcgcacccgtggttcgcagcaacaagaatagagactactgttcgaaggaacgggtgctt 180
 Db 287 gagattccagaggaacgagcaggaagattcgaagtagtcttcaataaaagaaatcgcgagatg 346
 Qy 181 ctgcagcagcaggtcccgactcgtcgtgagtcgaagggccagaggtggcccaacgatt 240
 Db 347 ctcaatgaggttctctgtgttgaatggttcttaataaaagaaatcgcgagatggt 406
 Qy 241 gctgaggaacctgatgaactccgctggaagacccagcgcgagataccgaagatcgcttgta 300
 Db 407 gctcgtctctcgtcgtcgtgaaattgacagcgtcgtgagatttccagat----- 458
 Qy 301 cagcagcgtcgtggtggaatgacaagatggcgcgtgaaatccgttcccatctccat 360
 Db 459 -atcaatctgtgaataagtttaaaaaattcaaggagaggttcgttcctcctcgat 517
 Qy 361 cacaattggcagcttgaaggtgctgctgagtcgagagccagcgcgagatcgacaatc 420
 Db 518 agaccatggcagattcaattgacggaggcaattgacgagggaacccgagatcgagc 577
 Qy 421 ctctgagatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
 Db 578 atctgggtctatggtccttatggttaaggggttaatacaacatgagagatggaagtcacta 637
 Qy 481 ctcaagcccgactggttctcacatggtggtggaacacgaagggacgtattgtaccagtac 540
 Db 638 aagaagattggttctcacaccgggtgggaagagagagagagagagagagagagagag 697
 Qy 541 atcgaggaccccaaaacgaaatttaactcctcgtatgaccaggtgtaatttaagattat 600
 Db 698 gacgaaggatctgacaagcatatgattggtgataattcctcgttgaatcagagattat 757
 Qy 601 aattatgacctgttagaattgttgaagacagagggaattcagtcgagacaaatcacga 660
 Db 758 aattatgagtaagagcattaaagagagagagagagagagagagagagagagagagag 817
 Qy 661 cttagttatcttggttcgacatgctgagatgactcgtatgctatgccaatgctcgtg 720
 Db 818 ataaagatagttgaattagtaaaatcacatgtaactgtaactgctgagagagagagag 877
 Qy 721 tatttgaataatcagcag 769
 Db 878 tctgtaaaatctccgaagatcgaaataaaatcatttattgtctgaagaa 926

QY	1	gttgggacgaggttg	caccttcaactggcgaacgacacccccaagattcatccacttg	60
Db	178	gtgtgcggcgaacaaactg	caactactggacagaaacacccctccgggattgtatcgcttc	237
QY	61	aagcgcggtcggcgaactg	caaggattgaagactgtctctgggaatgacagattccacttg	120
Db	238	aagacacaaaattcgtct	gtgtggttggaagaagaaatttggtaatc--gagctcaactgg	294
QY	121	gagccgaccctggttccg	acgaacagataagagactactgttcgaaggaaacgggtgctt	180
Db	295	gaaattcggagggcagcgt	ctcctcagatcgcgatatttgtctaaagaagaaacccctaat	354

Boevink PC, Ch

Key	Location/Qualifiers
PH	48..905
FT	/**tag= a
FT	
XX	
XX	
XX	WO9606932-A1.
XX	
XX	07-MAR-1996.
XX	
XX	
XX	30-AUG-1995; 95WO-AU00552.
XX	
XX	
PPR	07-NOV-1994; 94AU-0009281.
PPR	30-AUG-1994; 94AU-0007770.
XX	
XX	
PPA	(CSIR) COMMONWEALTH SCI & IND RES ORG.
PPA	(AUSU) UNIV AUSTRALIAN NAT.
XX	
PI	Boevink PC, Chu PWG, Keese PK, Khan RI, Larkin PJ;

WO9915646-A1.

XX
XX

WO9915646-A1.

Search completed: October 24, 2001, 10:07:20
Job time: 625 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 09:56:55 ; Search time 393.52 Seconds
(without alignments)
2059.923 Million cell updates/sec

Title: US-09-462-955-1
Perfect score: 1291

Sequence: 1 cgcacaaacacctctgtaag.....ggacggctgattgatctgg 1291

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179.2	13.9	1096	20	AAV71833
2	179.2	13.9	1096	21	AAV38950
3	163.4	12.7	1106	20	AAV71832
4	161.8	12.5	1106	21	AAV38946
5	159.6	12.4	1091	20	AAV71834
6	159.6	12.4	1091	21	AAV38947
7	145.4	11.3	886	20	AAV34687
8	112.4	8.7	1022	17	AAT13161
9	99.2	7.7	1017	17	AAT13165
10	81.2	6.3	593	20	AAV34686
11	62.8	4.9	1110	19	AAV24089

12	59.6	4.6	982	18	AAV49405	Banana bunchy top
13	59.6	4.6	1110	19	AAV24086	Banana bunchy top
14	59.6	4.6	1111	19	AAV24077	Banana bunchy top
15	59.6	4.6	1111	19	AAV24084	Banana bunchy top
16	59.6	4.6	1111	19	AAV24087	Banana bunchy top
17	58.6	4.5	1103	19	AAV24091	Banana bunchy top
18	58.6	4.5	1105	19	AAV24093	Banana bunchy top
19	58.6	4.5	1109	19	AAV24088	Banana bunchy top
20	58	4.5	1111	19	AAV24090	Banana bunchy top
21	57	4.4	1104	19	AAV24092	Banana bunchy top
22	56.4	4.4	1111	19	AAV24085	Banana bunchy top
23	49.4	3.8	936	22	AAF58252	Oligonucleotide D1
24	49.4	3.8	936	22	AAF58252	Oligonucleotide D1
25	49.4	3.8	936	22	AAF58254	Oligonucleotide D1
26	49.4	3.8	936	22	AAF58254	Oligonucleotide D1
27	49.4	3.8	936	22	AAF58257	Oligonucleotide D1
28	49.4	3.8	936	22	AAF58257	Oligonucleotide D1
29	49.4	3.8	936	22	AAF58259	Oligonucleotide D1
30	49.4	3.8	936	22	AAF58259	Oligonucleotide D2
31	49.4	3.8	936	22	AAF58262	Oligonucleotide D2
32	49.4	3.8	936	22	AAF58262	Oligonucleotide D2
33	49.4	3.8	938	22	AAF58255	Oligonucleotide D1
34	49.4	3.8	938	22	AAF58255	Oligonucleotide D1
35	46.4	3.6	287	20	AAV71830	Subgenomic fragmen
36	46.4	3.6	287	21	AAV38948	Banana bunchy top
37	46.4	3.6	300	20	AAV71831	Subgenomic fragmen
38	46.4	3.6	300	21	AAV38949	Banana bunchy top
39	38	2.9	244	22	AAF58238	Oligonucleotide D1
40	38	2.9	244	22	AAF58238	Oligonucleotide D1
41	36.4	2.8	1939	21	AAV29647	Human choline/etha
42	33	2.6	9345	21	AAZ93966	Mouse uromodulin p
43	32.8	2.5	1768	20	AAV83754	Porcine circovirus
44	32.6	2.5	5053	19	AAV22816	Nucleotide sequenc
45	32.6	2.5	5055	19	AAV17236	DNA from a region

ALIGNMENTS

RESULT 1
ID AAV71833 standard; DNA; 1096 BP.
AC AAV71833;
XX
XX 10-FEB-1999 (first entry)
XX BBTV DNA II clone (2-17) nucleotide sequence.
DE Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
XX Banana bunchy top virus.
XX
XX
FH Key Location/Qualifiers
FT TATA_signal 1..7
FT /*tag= a
FT stem_loop 8..38
FT /*tag= b
FT CDS 70..928
FT /*tag= c
FT /*product= "ORF-V2 product"
FT /*transl_except= (pos:215..217, aa:Gly)
FT polyA_signal 533..538
FT /*tag= d
FT polyA_signal 799..804
FT /*tag= e
FT polyA_signal 907..912
FT /*tag= f
FT polyA_signal 1030..1035
FT /*tag= ge
XX
XX US5846705-A.

[illegible]

RESULT 3
AAV71832
ID AAV71
XX

Db 690 cctatgtggagcaagatgacagacatagatttgatattcctccttgtaatacagg 749
 QY 803 agtatttaattatccctgttgaataatgtttaagacagggcaattcagttcgacaaat 862
 Db 750 attatttaattatgatataatagagcattaaaggatagggttatagagagactaaat 809
 QY 863 acgaaccccttagttatcttgggttcgacacatgtgcattgactcgtatttgccaatgtcc 922
 Db 810 acaaacccataaagatagttgaattagttgtaaaatcacatgtaacgtcgtcgaatttca 869
 QY 923 tgcctgattattgaaatcacagcagggacagagaataaaactgtggaatatttaagta 979
 Db 870 tgcctgactctgtaaaatctccgaagatogaataaaaatcattttattgtggaaga 926

RESULT 5

AAV71834

ID AAV71834 standard; DNA; 1091 BP.

XX

AC AAV71834;

XX

DT 10-FEB-1999 (first entry)

DE BBTV DNA II clone (2) nucleotide sequence.

XX

KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;

XX Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.

OS Banana bunchy top virus.

XX

PN US5846705-A.

XX

PD 08-DEC-1998.

XX

PF 06-APR-1995; 9505-0418071.

XX

PR 06-APR-1995; 9505-0418071.

XX

PA (BIOT-) DEV CENT BIOTECHNOLOGY.

XX

PI Soong T, Wu R, You L;

XX

DR WPI; 1999-059037/05.

XX

PT Nucleic acids having banana bunchy top virus component sequences -

XX used to design primers for use in polymerase chain reaction

XX detection of the virus

XX

PS Disclosure; Fig 12A-C; 27pp; English.

XX

CC This represents the nucleotide sequence of a banana bunchy top virus

CC (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention

CC provides nucleic acid sequences associated with BBTV that can be used in

CC a PCR technique for detecting BBTV. The nucleic acid sequences (AAV71830

CC to AAV71833) are used as the basis for the construction of PCR primers,

CC to detect BBTV infection. The PCR technique is used for detecting BBTV

CC in plant tissues (preferably banana, especially Musa species). The

CC virus, one of the most important banana species viruses, causes phloem

CC damage and is transmitted by aphids. PCR detection gives accurate,

CC reliable and specific determination of absence or presence of the

CC virus.

XX

SQ Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other;

Query Match

Best Local Similarity 12.4%; Score 159.6; DB 20; Length 1091;

Matches 477; Conservative 0; Mismatches 429; Indels 20; Gaps 4;

QY 66 cggctcccaactctgtcaaccgcgttggtcgtataaaagggttcctccatccgcgctg 125
 Db 28 cggctccctcgtcttctcctcgcgccctgacgtcatcattatgctcttttaagt 87

QY 126 gtgcttactctgaactacagacagagagcggcggaatgttgtaagcgcatcga 185
 Db 88 gtgcttactctgaatattattctcgcagcgagagagactttctcgtctcttgaa 147
 QY 186 gtcccttaactgtctatagttgggacagaggttgcaccttcaactgccaacg 245
 Db 148 ggagaaagattaactacgtctcgcgcagcagagtcgctccgagaccggtcgaa 207
 QY 246 acactccaaagattatccacttgaagacgggtcgcgcgactcaaggttggaagactgt 305
 Db 208 gcaactccaggatattctatccctgaagaaattattaaactgttggttggaagagag 267
 QY 306 tcttgggaatgacaggattcaactcgggacccggttctccgacgaacagaatagaga 365
 Db 268 gtactc---ttcgaaggtcactgggagagggcgcagaggaactgatgaacagaatcgag 324
 QY 366 ctactgttcgaaggaacgggtgcttctcgcagcagcaggtcccgactcgtccctggaatcaa 425
 Db 325 atactgttcgaagga---acccttgaaactgggttaactccggtggttccctggttcgaa 379
 QY 426 aaggccacgattggcccaacgatttctgaggaaacctgatgaactcgcctcgaaaccc 485
 Db 380 gaagcgaagctctcgtatagattcagagagagccctgaggaattgaagatggacgatcc 439
 QY 486 agcggatcacgaagatcggttgacacagagcttcggtggaatggacaagatgggcgcg 545
 Db 440 atcaaatctcgagatgcttggcagtggaatcaatataagatgccagaaatlaattccga 499
 QY 546 tgaataatccgttccatttccatlatcacaaattggcagcttgaaagtgtctgtcgcgcgtg 605
 Db 500 atgggttcacgaacta-----aaagaatggcaaaaataaataatcaacacatcga 550
 QY 606 agagccagcggacgacgcgaacatcctctggatctgcgcagcagcagggagagcgga 665
 Db 551 aggtgttctcgtatgcgaagatcatctgggtatcacgtctcctgcgcgagcgcaaggaaa 610
 QY 666 gtccgtgtttgccaaatatactcggactcaagcccaactgtttctacacatgtggtggaaac 725
 Db 611 gtcaacctcgcaagatatactatcaataaaacctggatgggatatacaacggtggaaa 670
 QY 726 cagaagggacgtattgtaccagtcacatcgagaccacaaacgaaatttaactctcgtgt 785
 Db 671 gacgtcggatattgatgcacatcataacgatggatcctgataatcattgattgatat 730
 QY 786 acccaggtgataattagagattttaaatattcctcgttagaatgtgttaagaacacagggc 845
 Db 731 cccgaagatcattcagattatctgaattatgctggttatagacaataatagaatagagt 790
 QY 846 attcagttcggacaaatcacgaacc---ccctagttatcttgggttcgaccatgtcagt 902
 Db 791 tttataataacaaatacgaaccatgtgtgattagaaaagatggacaaaatgtccatgt 850
 QY 903 actcgtatttgcaatgctcctcgtgatttattgaaataatcagcaggggacagaaataaact 962
 Db 851 aattgtattggcaaatgtgtcctgattattgtataaaatlttcagaagatagataaaaat 910
 QY 963 gtggaatatttaagatgtgtgcatc 988
 Db 911 aattaattgttagaaagaaacttc 936

RESULT 6

AAA38947

ID AAA38947 standard; DNA; 1091 BP.

XX

AC AAA38947;

XX

DT 23-AUG-2000 (first entry)

XX

DE Banana bunchy top virus related nucleotide sequence #2.

XX

KW Banana bunchy top virus; BBTV; detection; ds.

XX


```

XX
PR 07-NOV-1994; 94AU-0009281.
PR 30-AUG-1994; 94AU-0007770.
XX
XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX Boevink PC, Chu PWG, Keese PK, Khan RI, Larkin PJ;
PI Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
XX
XX WPI; 1996-160363/16.
DR
XX
XX Circovirus transcription regulatory sequences and related constructs
PT - useful in plants, esp. leguminous plants, for the modulation of
PT gene expression
XX
XX
FS Claim 4; Page 58; 121pp; English.
XX
XX The subterranean clover stunt virus (SCSV) genome has at least 7
CC distinct ssDNA components, designated segments 1-7 (AAL13160-66), each
CC contg. 1 major open reading frame and a non-coding region. Segment
CC 2 is predicted to be a viral replication-associated protein gene.
CC Genetic constructs useful in the genetic engineering of plants (esp.
CC legumes) comprise 1 or more heterologous gene(s) operatively linked
CC to a promoter region, and in some cases also a terminator region,
CC selected from segments 1-7. The transcription regulators facilitate
CC expression of foreign genes in plants and also facilitate control of
CC levels of gene expression in different plant tissue types.
XX
XX Sequence 1022 BP; 309 A; 183 C; 245 G; 285 T; 0 other;
SQ
Query Match 8.7%; Score 112.4; DB 17; Length 1022;
Best Local Similarity 49.2%; Pred. No. 7.1e-28;
Matches 423; Conservative 0; Mismatches 416; Indels 21; Gaps
QY 117 tcgccctggtcttcactctgaactacgacagacagagaagagcgcgcaattgttgaag 176
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 177 ggcatacgcagtccttaacttggctctgtatgttgggaacaggttcaccttcac 236
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 147 cctctctctcaagacgaattaaactattcgttgcgcgacgaactgcaact---ac 203
QY 237 tggcacaacgacacctccaagattaccacttgaagacacgctgcgcgactcaagatt 296
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 204 tggacagaacacctcccgaggattgtatcgttcaacaaattcgtcttgggtatt 263
QY 297 gaagactgtcttgggaatgacagattcacctggagccgacccgtggttcgcagca 356
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 264 gaagaagaatttgtaact--gagctcactggaaattgcgagagcagcgattctca 320
QY 357 gaatagagactactgttcgaaggacgggtgctcttcgcgacgagctccgactcgctc 416
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 321 gaatcgcatattgctgtaaagaacccctaattctgaaattggattccggtcatgaa 380
QY 417 tggagtcaaaagccacgattggcccaacgatttctgagaacctgatgaactcgct 476
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 381 gggttcgacaagcggagaacgataggagatttatgaagggatcccgaagaatgcaatt 440
QY 477 ggaagaccagcggaataccgaagatcggtttacacgagctcgttggaatggacaag 536
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 441 gaaggtccagatactgctcttcgatgaaggcgagaagaattgaaagagaattgttc 500
QY 537 atgggcccgtgaaatccgttccattccatatacaattggcagcttgaaagtctgtc 596
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 501 cgttatgatttcagaactcgt-----ccatggcaattgagcttcacga 548
QY 597 tgcgatcgagacgcagcgacgatcgacaactctctlggatatcgacgagacgagag 656
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 549 ggaatttaatgcggaacacgatgacgagatcaatctcgtgggtctatggttcagcgag 608
QY 657 agacgggaagtcctggtttgccaaatatctcgactcaagccgactggtttcacacatg 716

```


XX
PN US5756708-A.

US5756708-A.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 09:56:55 ; Search time 5701.85 Seconds
(without alignments)
3502.173 Million cell updates/sec

Title: US-09-462-955-1
Perfect score: 1291
Sequence: 1 ccaccaaaacctgctgaag.....ggacggcgtgagttgatctgg 1291

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_p11:
13: gb_p12:
14: gb_p13:
15: gb_p14:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_other:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_ro:
49: em_sts:
50: em_sy:
51: em_un:
52: em_vi:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_vil:
59: gb_vil2:
60: gb_vil3:
61: gb_vil4:
62: gb_vil5:
63: gb_vil6:
64: gb_vil7:
65: gb_vil8:
66: gb_vil9:
67: gb_vil10:
68: gb_vil11:
69: gb_vil12:
70: gb_vil13:
71: gb_vil14:
72: gb_vil15:
73: gb_vil16:
74: gb_vil17:
75: gb_vil18:
76: gb_vil19:
77: gb_vil20:
78: gb_vil21:
79: gb_vil22:
80: gb_vil23:
81: gb_vil24:
82: gb_vil25:
83: gb_vil26:
84: gb_vil27:
85: gb_vil28:
86: gb_vil29:
87: gb_vil30:
88: gb_vil31:
89: gb_vil32:
90: gb_vil33:
91: gb_vil34:
92: gb_vil35:
93: gb_vil36:
94: gb_vil37:
95: gb_vil38:
96: gb_vil39:
97: gb_vil40:
98: em_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1291	100.0	1291	58	CFDCG
2	179.2	13.9	1096	9	AR063452
3	179.2	13.9	1096	58	BYTV2
4	173	13.4	1095	58	AF216222
5	163.4	12.7	1106	9	AR063451
6	163.4	12.7	1106	58	BYTV1
7	160.8	12.5	1111	58	BBU12586
8	160.4	12.4	1127	58	BBU12587

M29963 Coconut fol
AR063452 Sequence
L32167 Banana bunc
AF216222 Banana bu
AR063451 Sequence
L32166 Banana bunc
U12586 Banana bunc
U12587 Banana bunc

Db 361 ACAGACTACTGTTCCGAGCAACGGGTGCTTCTCGACGACGAGTCCCGACTCGCTCTGGA 420
Qy 421 gTcaaaagccacgattggtcccaacgattgctgaggaacctgatgaactccgctctggaa 480
Db 421 GTCAAAAGCCACGATGCGCCCAACGATTTGCTGAGGAACCTGATGAATCCGCCCTGGAA 480
Qy 481 gaccagcggtatccacgaagatgctgtacacgaggttcggtgaatggacaagatgg 540
Db 481 GACCCAGCGCATACCGAAGATGGTGTACACGAGCTTCGGTGGATGGACAGATGG 540
Qy 541 gccctgaaatccgttccattccattccatcaaatggcagcttgaagtgtctgctgg 600
Db 541 GCCCTGAAATCCGTTCCCATTTCCATATCAAAATGGCAGCTTCAAGTGTCTGCTGG 600
Qy 601 atcgagagccagcgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 660
Db 601 ATCGAGAGCCAGCGGACGATGCGCAATATCTCGGACTCAAGCCCGACTGGTCTTACATGCTGG 660
Qy 661 ggaagtcggtgttgcgaatatctcgactcaagcccgactggtttcacacatggt 720
Db 661 GGAAGTCCGCTGTGTGCAAAATATCTCGGACTCAAGCCCGACTGGTCTTACATGCTGG 720
Qy 721 ggaaccagaagacgattgtaccagatcagacgacgacgacgacgacgacgacgacgac 780
Db 721 GGAACGAGAAAGACGATGTTACAGTACATCGAGGACCCCAAAACGAAATTTAATCTCTC 780
Qy 781 gatgacccaggtgtaattagtagtattaaattatgcctcttagaagtgtgtaagac 840
Db 781 GATGTACCCAGGTGTAATTTAGAGTATTTAAATTTATGCCCTGTTAGATGTTTAAAGAC 840
Qy 841 agggcattcagtcgacgaataacgaaccccttagtattcttggttgcacattgtgcat 900
Db 841 AGGGCATTCAGTTCGACGAATAACGAACCCCTTAGTATTCTTGGGTTTCGACATGTGCAT 900
Qy 901 gtactcgtattgccaatgctcgtcgtgattatttgaataatcagcagagcagataaaaa 960
Db 901 GTACTCGTATTTGCCAATGCTGCTGATATTTGAAATTCAGAGGAGCAGAAATAAA 960
Qy 961 ctgtgaatatataagtagtctgcatctaaattacacccatgacccgcccgcacacgc 1020
Db 961 CTGTGGAATTTTAAAGTATGTGTCTATCTAAATTTACACCAATACCCGCCGCCACGCC 1020
Qy 1021 tatcgtttacatttatgaataatcctgcccagcgacgagcgtggaggtgctaccgcgc 1080
Db 1021 TATCGTTTACATCTTATGAAATATCTGCCCCAGCGCCGAGGCTGGAGGTGCTACCCGCC 1080
Qy 1081 cgaagccggggaacaatatgaatcaggttatggcgggccacacataaaagattccatt 1140
Db 1081 CGAAGCCGGGAACAATATGAATCGAGTTATGGCGGGGCCACCAATAAAAGATTCCATTT 1140
Qy 1141 ggataagaacgaatctgttacttctgttgagtgagtcgacgacgacacacacacacac 1200
Db 1141 GGATGAAGAACGAATCTGTACTTTGCTGTGAGTGCAGCGCAACCACTTCCACGTCACCA 1200
Qy 1201 tccaggtgagtagctgtgtagaagaagcgcgaagcgcgaagcgcgaagcgcgaagcgc 1260
Db 1201 TCCAGGTGAGTAGCTGTCTGAGAGAAAGCCGCAACCCCTCTATACCGTCCATTGTT 1260
Qy 1261 ttgcgcatcgagcgtgagttgactggtg 1291
Db 1261 TTTGGCGATCGGCGGCTGAGTTGATCTGG 1291

RESULT 2
AR063452 1096 bp DNA PAT 29-SEP-1999
LOCUS AR063452
DEFINITION Sequence 4 from patent US 5846705.
ACCESSION AR063452
VERSION AR063452.1 GI:5992760
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

Unclassified.
1 (bases 1 to 1096)
Wu, R., You, L. and Soong, T.
Nucleotide sequence of two circular SSDNA associated with banana
bunchy top virus and method for detection of banana bunchy top
virus
Patent: US 5846705-A 4 08-DEC-1998;
JOURNAL Location/Qualifiers
FEATURES 1..1096
BASE COUNT 347 a 231 c 244 g 274 t
ORIGIN
Query Match 13.9%; Score 179.2; DB 9; Length 1096;
Best Local Similarity 52.4%; Pred. No. 1.6e-42;
Matches 499; Conservative 0; Mismatches 433; Indels 20; Gaps 4;
Qy 40 agccgcggttagtattaccaccccggtcccaacccctgtgtaaccccgcttggcta 99
Db 7 AGGAGGAGCGGCTAGTATTACCGCTCTCTCGCACCTTCTCTCGCACCTGACGTC-- 64
Qy 100 taaatgggttcctccattccgctgcttcaactctgaactcagacagaggaag 159
Db 65 ---ATCATTTATGTCCTCTTTAAATGGTGTCTTCACTCTGAATATTCTCTCGCGACGGAG 121
Qy 160 ggcgcgaatgttgaaggcgcatcgatcccttaacttggctctatgtctatgttgggac 219
Db 122 CGAAGAAGACTTCTCGCTCTTCTGAAGGAGGAGGATGTTCACTACTCTCTCTCGCGCAG 181
Qy 220 gaggttgcaccttcaactgccaacgacacaccccaagattccacttgaagaccggt 279
Db 182 GAAGTCTCTCGCGCCACCGCCGAGAGCAACCTCCAGGATATCTATCTCTGAAAAATCA 241
Qy 280 cggcgactgcaaggattgaagactgttctgggaatgacagattcactcagaccgacc 339
Db 242 ATTGCGCTCGCGGATTTGAAAAGAAAGTATGG--CTCCCGTGTCTACTGGGAGATTGCG 298
Qy 340 cgtgttcgcagcaagaatagagactactgttcgaaggacgggtgcttctcgcagcac 399
Db 299 AAAGGAAGTGACGACACAGATCGCAGATCTGTTTGAAGGAAACCTAGTCTTGAACCTG 358
Qy 400 ggaagtcgcgactcgtcctggagtcacaaaggccacgattggcccaacgatttctgagaa 459
Db 359 GGTACTCCGGTGTCTCTGTTGTTGAAGGAGCGCAAGCTTCTCGATAGATTACAGAGAGC 418
Qy 460 cctgatgaactccgctggaagacccagcggtatcaccgaagatgctgttacacgaagct 519
Db 419 CCTGAGGAATTAAGATGGACGATCCATCCAAAGTATCGCAGATGCTTGGCAGTGGATCA 478
Qy 520 tcggtggaatgacaaagatggcgctgaaaaatcgttcccatctccatacacaattgg 579
Db 479 ATTAAGATGCCAGATTAATTCGAAATGGGTTTCACCAACTA-----AAAGAATGG 529
Qy 580 cagcttgaagtgtctgtcgtgagagccgagccgagcagcagcagcagcagcagcagcagc 639
Db 530 CAAATAAATTAATTCACACATCGAAGTGTCTCTGATGATGATGATGATGATGATGATGAT 589
Qy 640 tgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 699
Db 590 TACGGTCCCAACGAGGCGGAGGAAAGTCAACCTTCGAAGATATCTATCATTTAAACCC 649
Qy 700 gactggttctacacatgtgtggaacccagaaagacgattgtaccagtcacatcagggac 759
Db 650 GGATGGGATATATCAACCGGTGGAAAGACGTCGGATATATGATGATGATGATGATGATGAT 709
Qy 760 ccaaaacgaatattaactcctcgtatcaccaggtgtaatttagagatttatttattatgcc 819
Db 710 CCGATAAATCATTCGATATTGATATATCCAGAGTCAATTCAGATTATCTGAATTATGCG 769
Qy 820 ctgttagaatgtttaagaacacagggcattcagttcggacacaaacagacacacacacac 876
Db 770 GTTATAGACAAATTAAGATAGAGTTTAAATAAATAACAAATACGAACCATGCTGATT 829


```

AF216222      1095 bp      DNA      VRL      08-MAR-2001
LOCUS          Banana bunchy top virus satellite S2 replication initiation protein
DEFINITION    (ORF V1) gene, complete cds.
ACCESSION     AF216222
VERSION       AF216222.1 GI:12004327
KEYWORDS
SOURCE        Banana bunchy top virus.
ORGANISM      Banana bunchy top virus.
REFERENCE     Viruses: ssDNA viruses: Circoviridae.
AUTHORS       1 (bases 1 to 1095).
TITLE         Horser, C.L., Karan, M., Harding, R.M. and Dale, J.L.
              Additional rep-encoding DNAs associated with banana bunchy top
              virus
JOURNAL       Arch. Virol. 146 (1), 71-86 (2001)
MEDLINE       21163975
REFERENCE     2 (bases 1 to 1095)
AUTHORS       Horser, C.L., Karan, M., Harding, R.M. and Dale, J.L.
TITLE         Direct Submission
JOURNAL       Submitted (14-DEC-1999) School of Life Sciences, Queensland
              University of Technology, GPO Box 2434, Brisbane, QLD 4001,
              Australia
FEATURES      Location/Qualifiers
               1..1095
               /organism="Banana bunchy top virus"
               /isolate="Taiwan"
               /db_xref="taxon:12585"
               /note="satellite S2"
               64..921
               /gene="ORF V1"
               64..921
               /gene="ORF V1"
               64..921
               /gene="ORF V1"
               /codon_start=1
               /product="replication initiation protein"
               /protein_id="AAG44004.1"
               /db_xref="GI:12004328"
               /translation="MSSFWCTLNYSAAERDFLALKEFLNAYVGVDEVPSTG
               RKHQGLSLKSIKIGLKKRYSKAHMERARGTDEQNRGYCKETLVLLELGTVPVP
               GSKRKLLEFRPEELKEMEDPSKYRCLAVESLNARKNSEWHELREWNKLIQH
               IFQVDPDRSIIWYVGNPGEKGSTFARYLSLKPWGVIINGKTSMDMLIIMDPDNHW
               IIDIIPRSHDLYNGVIEQIKRVLINIKYPCVIRKDGQNVHIVMANVLPDYCKIS
               EDRIKTIK"
BASE COUNT    347 a 219 c 252 g 277 t
ORIGIN
Query Match   13.4%; Score 173; DB 58; Length 1095;
Best Local Similarity 51.8%; Pred. No. 1.1e-40;
Matches 473; Conservative 0; Mismatches 425; Indels 15; Gaps 3;
QY 66 cggctcccaacctgtgtataaccccgcttgctataaatgggttcctccattgcgcgtg 125
DB 21 CCGCTCCTCCTCGCTTCTCCTCCTCCCTCGAGGTCATCATTTATGCTCTCTTAAATG 80
QY 126 gtacttcaactgaactagacagagagagagagagagagagagagagagagagagag 185
DB 81 GTGCTTCACTCTGAATTTATCTCTCCACGCGGAGGAGAGACTTCTCGCTCTTCTGAA 140
QY 186 gtoccttaactgtgtatgtatagttgggagagagagagagagagagagagagagag 245
DB 141 GGAGGAAGAGTTAAATAGCTGTGCTGGCGACGAGAGTCCGTCAGCAGCGGTCGAA 200
QY 246 acactccaaagattatccactgaagacgggtcgagactgcaaggattgaagactgt 305
DB 201 GCACCTCCAGGATATCTATCTCCTGAAGAAATCTATTAAAGCTGTGGATTGAAGAA 260
QY 306 tcttgggaatgacagagattcaactgagcccgctggttcccgacgacagacagataga 365
DB 261 GTACTC---TTCGAGGCTACTGGGAGGCGGAGAGAGACTGATGAACAGAAATCGCGG 317
QY 366 ctactgttcgaaggaacgggtgtcttctcgagcagcaggagtcctcgactgtctctgag 425

```

```

Db 318 ATACTGTTCCAGGAACCCCTAGTTCTTGAACGTGGGTACTCCGGTGGTCCCTGGTTCGAA 377
QY 426 aaggccacgattggcccaacgatttctgaggaaacctgtatgaactcgcctggaagacc 485
DB 378 GAAGCGCAAGCTTCTCGAGAGATTTCAGAGAGAGCCCTGAAGAATTTCAGATGAGGATCC 437
QY 486 agcgagataccgaagatggtgtacacagagagcttcggttgaatggaagaatggccgc 545
DB 438 ATCAAAGTATCGCAGATGCTTGGCAGTGAATCGTTGAACAATGCTAGGAAAAATTTCTGA 497
QY 546 tgaatatccgttccattccatatacacatggcagcttgaagctgtcgtcgcacatcgg 605
DB 498 ATGGGTTTCACTAAGAGATTGGCAAAATAAATAATTAATTAAC-----ACATCGA 548
QY 606 agagcagcggacgacgcacacatcctctggtatgcgacgagacgagcggagcggaa 665
DB 549 AGGTGTTCTGTATGATCGAAGTATCATCTCGGTATACGGTCCCAACGGAGGCCAAGGAA 608
QY 666 gtccgtgttgcacaaatctcggactcaagccgactgtttctacacatgtggtggaac 725
DB 609 GTCAACCTTCGCAAGATATCTATCATTAACACCTGGATGGGATATATCAACCGTGGAA 668
QY 726 cagaaggacgtattgtaccagtcacatcgagagaccccaacaaacgaatttaatcctcgt 785
DB 669 GACGTCCGATATGATGCACATCATACGATGATCTGTATATCATTCATTTATGATAT 728
QY 786 acccaggttaatttagagtatttaaatcctcgtttagaatgttgaagacagcggc 845
DB 729 CCCCAAGTTCATTCAGATTATCTGAATTTATGGCGTTATAGAACAATAAAGATAGAGT 788
QY 846 attcagttcggacaaatcacgaacc---ccttagttattcttgggttcgacatgtcagt 902
DB 789 TTTAATAATACAAATACGAACCATGTGTGATTAGAAAAGATGCAAAATGTCCATGT 848
QY 903 actcgtattgccaatgtcctcctgattatttgaataacagcagcagcagcagcagcagc 962
DB 849 AATTGTTATGGCAATGTGTTCCCTGATTATTGTAAATTTTCAAGAGATAGATAAATAAT 908
QY 963 gtggaatatltaa 975
DB 909 AATTAATTGTTGA 921
RESULT 5
AR063451
LOCUS          AR063451      1106 bp      DNA      PAT      29-SEP-1999
DEFINITION    Sequence 3 from patent US 5846705.
ACCESSION     AR063451
VERSION       AR063451.1 GI:5992759
KEYWORDS
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     1 (bases 1 to 1106)
AUTHORS       Wu, R., You, L., and Soong, T.
TITLE         Nucleotide sequence of two circular ssDNA associated with banana
              bunchy top virus and method for detection of banana bunchy top
              virus
JOURNAL       Patent: US 5846705-A 3 08-DEC-1998;
FEATURES      Location/Qualifiers
               1..1106
               /organism="unknown"
BASE COUNT    335 a 225 c 257 g 289 t
ORIGIN
Query Match   12.7%; Score 163.4; DB 9; Length 1106;
Best Local Similarity 50.6%; Pred. No. 8.9e-38;
Matches 454; Conservative 0; Mismatches 431; Indels 12; Gaps 2;
QY 83 ctaaccgcgttggtataaatgggttcctccattcccgctggtcctcactctgaact 142

```


Db 739 GATATTCCTCGCTGTAATCAGGATTTTAAATTTATGATTTATAGAGGATTAAGGAT 798
Qy 841 aggcattcagttcggacaatacgaaccccttagttatcttgggttcgaccatgtgcat 900
Db 799 AGGGTGATAGAGAGTACTATAAATTAACCTATTAAAGTTAGTTGAATTAATATACAT 858
Qy 901 gtaactcgtatttgcaatgctcctcgtcgtgatttattgaaaatcagcagggacagaataaaa 960
Db 859 GTAATTGTAATGGCTAAATTCATCCAGAAATCTGTAAATCTCCGAAGATAGAATAAAG 918
Qy 961 ctgtggaatatttaa 976
Db 919 ATTATTTATTGTTAAA 934

RESULT 8
BBU12587 1127 bp DNA circular VRL 01-FEB-1995
LOCUS Banana bunchy top virus DNA IV ORF V1 and ORF C1 genes, complete
DEFINITION cds.
ACCESSION U12587
VERSION U12587.1 GI:642393
KEYWORDS Banana bunchy top virus.
SOURCE Banana bunchy top virus.
ORGANISM Viruses; ssDNA viruses; Circoviridae.
REFERENCE 1 (bases 1 to 1127)
AUTHORS Wu, R.-Y. and You, L.-R.
TITLE Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related viruses
DNAs
Unpublished
REFERENCE 2 (bases 1 to 1127)
AUTHORS Wu, R.-Y.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1994) Rey-Yuh Wu, Agricultural Biotechnology Division, Development Center for Biotechnology, 81 Chang Hsing Street, Taipei, 10671, Taiwan, ROC
FEATURES
source
1. .1127
/organism="Banana bunchy top virus"
/db_xref="taxon:12585"
/clone="S15 and 4-2"
TATA_signal
stem_loop
CDS
1. .8
15. .49
79. .933
/note="ORF V1"
/codon_start=1
/protein_id="AAA61877.1"
/db_xref="GI:642394"
/translation="MSSPKWCFTLNYSAAEREDFLALLKEELNYAVVGDEVAPSSG QKHLQYLSLKKSLKGLKKKYSRAHWEARGSDENAKYCSKETLLELGFPSAQ GSNRKLSEMSRSRPERMKIEQPEIYHRYTSYKLLKKEFEVHPCLDRPWOIQLTEA IDEEPDRTIFWVYCPNENEGKSTYAKSLMKDMFTRGGKKNLISYVDEGSEKHI VFDIPRCNDYLDNYVIALKDRVIESTKYRPIKLVELNLIHVIMANFMFECKISE DRKIITYC"
complement(497..643)
/note="ORF C1"
/codon_start=1
/protein_id="AAA61878.1"
/db_xref="GI:642395"
/translation="MTSHMLISPHYSDHPRRSFDHLVPRQLPPSIESAMVYRGKDE RIPP"
BASE COUNT 353 a 222 c 260 g 292 t
ORIGIN

Query Match 12.4%; Score 160.4; DB 58; Length 1127;
Best Local Similarity 52.3%; Pred. No. 7.1e-37;
Matches 459; Conservative 0; Mismatches 401; Indels 18; Gaps 4;
Qy 124 tgggtcttactctgaactacagacagagagagcgcgcaatcttgaaggcgcatc 183
|||||

Db 94 TGTGCTTCACTCTGAACACTCTCTCCGAGCTGAGCGAGAACTTTTCGCTCTCTG 153
Qy 184 ggtcccttaacttggtctatctatagttgggagcagaggttgccacttcaactgcca 243
Db 154 AAGGAGGAAGAGCTTAACCTACGAGTGTTCGGCGAGGAAGTCCCTCCGAGCTCCGCCAG 213
Qy 244 cgacactccaaggattccacttggaagacggttcggcgactgcaaggattgaagact 303
Db 214 AAGCAGCTACAGGATATCTATCCCTGAAGAAATCTATAAACTTTGGAGGATGAAGAAG 273
Qy 304 gttcttgggaatgaacagattcaacctggagcagcccggttgcgacagacaagaataga 363
Db 274 AAGTAT---TCTTCGAGAGCTCACTGGGAGGGCGAGAGGATCTGACAGATAATGCT 330
Qy 364 gactactgttcaagaaacgggtgctctcgcagcagcgagtcgccgactcgtcctgagtc 423
Db 331 AAGTACTGTTCTTAAAGAAACCTTAATCTTGAATTTGGGGTTTCCGCCCTCTCAAGGCTCT 390
Qy 424 aaaagccacgattggcccaacgatttgcctggaggaacctgatgaactccgcctcgtgaagac 483
Db 391 AATAGCGGAACACTATCGGAGATGGTTTCTAGATCTCCAGAACGCATGAGAAATTGAACAG 450
Qy 484 ccaggcggaatcacgaagatgctgtgtacacggaggttcggtggaatggacaagatgggcc 543
Db 451 CCTGAGATATACACAGT-----ACACATCTGTGAAGAAGTTAAAAAAATTCAG 501
Qy 544 gctgaaatccgttccattccatcaccaattggcagcttgaagtgtgtcgtcgcac 603
Db 502 GAGGAATTCGTTTCATCTTCCCTCGATAGACCATGCGAGATTCATTCACGGAGGCAATT 561
Qy 604 ggagccagcggacatgcacaaatcctctgatatgagcagcagcagcagcagcagcgg 663
Db 562 GACGAGGAACAGATGATCGAAGCATCTTCTGGGTCTATGGTCCGAATGTAATGAGGGG 621
Qy 664 aagtcggttggccaaatctcgcgactcaaccccgactggttcacacatggtgga 723
Db 622 AAATCAACATATGGAAGTCAATT---AATGAAGAAGGACTGGTCTACACACAGGTTGG 678
Qy 724 accagaaggacgtattgtaccagtcacat---cgaggaccccaaacgaaatttaactc 780
Db 679 AAGAGGAGGAACATACCTGTTCTCTAGTGGAGCAAGGATCTGAGAACGATATGTTATTT 738
Qy 781 gatgacccaggtgtaatttagagatttataatgacctgttccctgttagaattgttaagaac 840
Db 739 GATATTCCTCGCTGTAATCAGGATTAATTAATATATGTTATAGAGGATTAAGGAT 798
Qy 841 aggcattcagttcggacaaatcgaaccccttagttatcttgggttcgaccatgtgcat 900
Db 799 AGGGTGATAGAGAGTACTATAAATTAACCTATTAAGTTAGTTGAATTAATATACAT 858
Qy 901 gtaactcgtatttgcaatgctcctcgtcgtgatttattgaaaatcagcagggacagaataaaa 960
Db 859 GTAATTGTCATGGCTAAATTTTCATGCCAGAAATCTGTAAATCTCCGAAGATAGAATAAAG 918
Qy 961 ctgtggaatatttaaagtatgtgcattcctaaattaac 998
Db 919 ATTATTTATTGTTAAATAACACGCTATGACAAATCGTAC 956

RESULT 9
AR063453 1091-bp DNA PAT 29-SEP-1999
LOCUS Sequence 5 from patent US 5846705.
DEFINITION AR063453
ACCESSION AR063453
VERSION AR063453.1 GI:5992761
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1091)
AUTHORS Wu, R., You, L., and Soong, T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top

JOURNAL Patent: US 5846705-A 5 08-DEC-1998;
 FEATURES Location/Qualifiers
 1. .1091
 source /organism= "unknown"
 BASE COUNT 343 a 224 c 248 g 276 t
 ORIGIN

Query Match 12.4%; Score 159.6; DB 9; Length 1091;
 Best Local Similarity 51.5%; Pred. No. 1.2e-36;
 Matches 477; Conservative 0; Mismatches 429; Indels 20; Gaps 4;

```

QY 66 cggctccccaacctctgctaaaccgcgcttggtcgtataaaatgggttctctcattcgccgctg 125
DB 28 CGCGTCTCTCTCGCTTTCTCTCTCGCCCTGACGTCATCATATGTCTCTTTTAAATG 87

QY 126 gtgcttcaactctgaatacgcagacagagagagcggcggaatgttgtaagcgcatcga 185
DB 88 GTGCTCTACTCTGAATATTCTCTCGCAGCGGAGGAGAACTTCTCGCTCTCTCTGAA 147

QY 186 gtcccttaacttggtctatgtatagttggggaagaggttgcaacttcaactgccaacg 245
DB 148 GGAGGAAGAGTTAAATTACGCTGTCTGCGCAGCAAGTCGCTCCGAGCACCGGTCGGAA 207

QY 246 acacctcaagagattccacttgaadaccgcgtggcgactgcgaagattgaagactgt 305
DB 208 GCACCTCCAGGGATATCTATCCCTGAAGAAATCTATTAAAGCTTGCTGATTGAAGAAGAG 267

QY 306 tcttggaatgacaggattcaccttgagcgcacccgtggttccagcaagaataagaga 365
DB 268 GTACTC---TTCGAAGCTCACTGGGAGGCGGAGAGAACTGATGAACAGAATCGCAG 324

QY 366 ctactgttcgaagaaacgggtgctctcgcagcagcgagtcgccactgctccgtgagtc 425
DB 325 ATACTGTTTGAAGGAA-----ACCCCTTGAACCTGGGTACTCCGGTGGTCTCTGGTTCGAA 379

QY 426 aaggccacgattggccaacgatttgtgagaaacctgatgaactccgcctggaagacc 485
DB 380 GAAGCGCAAGCTTCTCGATGATTACAGAGAGAGCCCTGAGGAATTTGAAGATGGACGATCC 439

QY 486 aggcggataccgaagatgcgtgttacacgagcttcggtggaatggacaagatgggcgc 545
DB 440 ATCCAAGTATCGCAGATGCTTGCGCAGTGGATCAATTAAGATGCCACAATTAATTCGA 499

QY 546 tgaataatcgtctccattccatcaacaattggcagcttgaagtgctgctgcgactcg 605
DB 500 ATGGGTTTCAAGACTA-----AAAGATGCAAAATAAATAATTAATCAACACATCGA 550

QY 606 agagccagcgacgatcgcaaatcctctgatatcgcgacgagacggaagagacgggaa 665
DB 551 AGGTGTTCCGTGATCGAGTATCATCTGGGTATACGGTCTCTGCCGAGGCGCAAGGAAA 610

QY 666 gtccggttttgccaaatattctcggactcaagccgcactggttctacacatggtggtgaac 725
DB 611 GTCAACCTTCGAAGATATCTATCATTTAAACCTGGATGGGGATATATCAACCGTGGAAA 670

QY 726 cagaaaggacgtatgtaccagtacatcgaggacccaacaaatttaactcctgatgt 785
DB 671 GAGCTCGGATATGTCACATCATACAGATGGATCTCTGATAATCAITGGATTATTGATAT 730

QY 786 acccagtgtaattagatgatttaattatgcctgttagaatgtgttaagaacagggc 845
DB 731 CCCAGAAGTCATTACAGATTATCTGAAATATGGCGTTATAGAACAAATTAAGAATAGAGT 790

QY 846 attcagtcggacaataacacgaacc---ccttagttatcttgggttcgacatgtgatgt 902
DB 791 TTTAATAAATAACAAATACGAACCATGTGTGATTAGAAAGATGGACAAATATGTCATGT 850

QY 903 actcgtatttgcaatgtcgtcgtgatttataattgaaatacagcaggagacagataaact 962
DB 851 AATTGTTATGGCAATGTGTGCTGATTATTCTGAAATTTTCTGAGANGATAGAAATAAAAT 910
  
```


JOURNAL		Submitted (04-OCT-1993) Yu-Chan Chao, Institute of Molecular Biology, 120, Sec. 2., Yen-joe-yuan Rd., Nan-Kang, Taipei, Taiwan	
FEATURES		Location/Qualifiers	
source		1. .1095	
		/organism="Banana bunchy top virus"	
		/specific_host="Musa acuminata Colla"	
		/db_xref="taxon:12585"	
		/clone="pBtd18"	
		/note="component 2 (BBTV-C2)"	
		82. .390	
		/note="orf1"	
		/codon_start=1	
		/protein_id="AA17783.1"	
		/db_xref="GI:475434"	
		/translation="MSSSLKWCFTLNYSAAERENFLSKKEEDVHVAVGVDEVAPA	
		TQKHLQGLYLSLKKMPRIEEVWLPCLSGDCQRRRRREFVLQRPNPFRIVSCC	
		"	
		419. .634	
		/note="orf2"	
		/codon_start=1	
		/protein_id="AA17784.1"	
		/db_xref="GI:475435"	
		/translation="MIARSPDRMKTEQPEIFHRYOSVNLKFKKEEFVHPCLDRPQWI	
		QLTEAIDEEPDDRSLIIVWYPYGNKGK"	
BASE COUNT		337 a	217 c 253 g 288 t
ORIGIN			
Query Match		11.1%; Score 143.6; DB 59; Length 1095;	
Best Local Similarity		49.9%; Pred. No. 8e-32;	
Matches		451; Conservative 0; Mismatches 439; Indels 14; Gaps 3;	
QY	95	ggctataaagggttctccattcgcgcgtggtgcttccactcgaactcgaacagagag	154
DB	74	gcctggagatgtccagtcctctcttaagtggtgcttcaactcgaattactctccgcgg	133
QY	155	aagagggcggaattgttaaggcgcatcgagtccttaacttggtctgtatgtatgtt	214
DB	134	cagagagagaaacttctctcttctgaaggaggagatgttcaactacgctgtctcgc	193
QY	215	gggacgggttgacacttcaactggccacacacacacacacacacacacacacac	274
DB	194	gcgacgaagtgcctccgcccacacacacacacacacacacacacacacacacac	253
QY	275	ccggtcgagactcaaggattgaagactgttcttggaatgacagattcacctggagc	334
DB	254	agaaatgcctcgcc--ggattgaagagagatgg--ctcccggtgctcactggaga	308
QY	335	cgaccgtgttcgacgacagacacacacacacacacacacacacacacacacac	394
DB	309	ttgccagaggaacacagacagagaaattcgaaagtactgttccaaagaaaccc	368
QY	395	agcaggaagtcgcactcgtctgaagccacacacacacacacacacacacacac	454
DB	369	aattaggggttccgtgttgaatggttcttaataaagaaatattcgagattgctc	428
QY	455	aggaacctgatgaactcgcgcgtggaagccacacacacacacacacacacacac	514
DB	429	gttctcctgatcgcatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa	479
QY	515	gagcttcggtgaatggacagatggccgctgaaatccggttccattccatcata	574
DB	480	aattctgtaataagttataaataaataaataaataaataaataaataaataaata	539
QY	575	attggcagcttgagctgtctgcgtcgagagccagcagcagcagcagcagcagc	634
DB	540	catgccacatacaattgacggagcaattgacggagcaattgacggagcaattgac	599
QY	635	ggatatcgacgacgagagagagagagagagagagagagagagagagagagag	694
DB	600	gggtctatggtccttattgtaaggggtaataaataaataaataaataaataaata	659

QY	244	cyacacctccagagattccattccattgaagccggtgcgcactgcgaagattgaag	303
DB	198	aagcaccatcagggatattctatccctgaagaaattctataaacttggagattga	257
QY	304	gttcttggaatgacagattcactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	363
DB	258	aagtatt--cttcgagagctcactgggagaggggagagatctgacgaatgact	314
QY	364	gactactgtcgaaggaacgggtgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	423
DB	315	aagtattgtcttaagaacaccttattcttgaattgggggttccgcgcctcaaggct	374
QY	424	aaagccacgattggccacagatttgcgaggaacacgtgatgaactcgcgcgcgcgc	483
DB	375	aatagacgaaactatcgagatggtttctagattccagaacgcattgagaaattgac	434
QY	484	ccagcggataccgaagatgctgtacacgagcttcgggtggaatgacaaatggg	543
DB	435	cctgagatattacagat-----acacatctggaagaaattgaaatgacaaatg	485
QY	544	gtcgaaatccgttccattccattccattccattccattccattccattccattcc	603
DB	486	gaggaattcgttcattccttgcctgcgatagaccatgcagattcaattgacggagc	545
QY	604	ggagcccgagcagatcgacacacacacacacacacacacacacacacacacac	663
DB	546	gacgagaaacccgatgatcgaaagcattctcgtggtctatgggtccgaattgga	605
QY	664	aagtcggttggcaaatatctcggaactcaagccgcgcgcgcgcgcgcgcgcgcgc	723
DB	606	aatcaacatattgcaagtcatt---aatgaagaggactggtcttacaccagaggtg	662
QY	724	accagaagagcagattgtaccagtcacat---cgaggaccccaacaaatgaa	780
DB	663	aagaagagaaactatgcttcttactgagcgaagagatctgagaagcatattgtat	722
QY	781	gatgaccaggtgtaattagatgatttaaatgacacacacacacacacacacac	840
DB	723	gattattcctcgctgtaatacaggaattttaaattatgattgattgagggcaatt	782
QY	841	agggcattcagtcgcaaatcagaccccttagttatcttggttgcacacattgcat	900
DB	783	aggggtatagagagactaaatataaacttattgaattgattgattgattgattg	842
QY	901	gtactcgatttgcgaatgctcgtcgtgatttattgaaatcagcagcagcagcaga	960
DB	843	gtaattgctatggtctaatctatgacagaaattctgtgaaatctccgaagata	902
QY	961	ctgtggaatttttaagatgtgtcatctaaattacacacacacacacacacacac	1012
DB	903	attattttttgtaaaagaaattattattatgacattacacacacacacacacac	954

RESULT	11
U02312	
LOCUS	U02312 1095 bp DNA VRL 28-APR-1994
DEFINITION	Banana bunchy top virus component 2.
ACCESSION	U02312
VERSION	U02312.1 GI:475433
KEYWORDS	
SOURCE	Banana bunchy top virus.
ORGANISM	Banana bunchy top virus.
REFERENCE	1 (bases 1 to 1095)
AUTHORS	Yeh, H.-H., Su, H.-J., and Chao, Y.
TITLE	Genome characterization and identification of viral-associated dsDNA component of banana bunchy top virus
JOURNAL	Virology 198, 645-652 (1994)
MEDLINE	94120739
REFERENCE	2 (bases 1 to 1095)
AUTHORS	Chao, Y.
TITLE	Direct Submission

VERSION AJ005966.1 GI:3550532
 KEYWORDS C9 gene; component 9; putative; rep protein; replication associated protein.
 SOURCE faba bean necrotic yellows virus.
 ORGANISM faba bean necrotic yellows virus.
 REFERENCE 1 (bases 1 to 1004)
 AUTHORS Katul, L.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
 REFERENCE 2 (bases 1 to 1004)
 AUTHORS Katul, L., Timchenko, T., Gronenborn, B. and Vetter, H.J.
 TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
 JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
 MEDLINE 99094637
 FEATURES Location/Qualifiers
 1..1004
 /organism="faba bean necrotic yellows virus"
 /viral
 /isolate="SV292-88"
 /db_xref="taxon:59817"
 73..918
 /gene="C9"
 73..918
 /gene="C9"
 /function="putative replication associated (rep) protein"
 /codon_start=1
 /product="component 9"
 /protein_id="CAA06789.1"
 /db_xref="GI:3550533"
 /db_xref="SPTREMBL:O91252"
 /translation="MSAVNNVFTLNFAAGEVPVLSFDERVOYAVQWHERVNHDIQCVI
 OLKKAKNTVKNITIGGNPHLEKMGKSGIEASAYAKKEESRVAGPWSYGGELKKGSHK
 RKIMELIKDPENLEPEQKYRRAMASAMDESRLAEFGFPYMFYSWQETVLGLLEE
 EPNDRTIIVYGPNGNEKSGQFGKFLGKDKDLYLPQGTQDMYMLMKNPKANVMD
 IPRCNSEYLNQFMELIKNRTIFSYKYPVGCILNNKHIVILANVLPDYKISQDRI
 KLIYC"
 BASE COUNT 352 a 159 c 226 g 267 t
 ORIGIN
 Query Match 10.2%; Score 132; DB 59; Length 1004;
 Best Local Similarity 54.0%; Pred. No. 2.5e-28;
 Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
 QY 476 tggagaccagcggatgacacgaagatgcgtgtgtacacgagcgttcggtggaatggacaa 535
 DB 419 TGGAGAACCCCAAGAAATACAGAGAGCTATGCTTGTGTCGCCCATGGACGAATCTCGGA 478
 QY 536 gatggccgctgaaatccggttccattccatatacacaattggcagctgaagtgcgt 595
 DB 479 AGCTTGGGGAAGAAAGAGGCTTCCCTATATGTTTACAGCTGGCAAGAAACAGTGTGG 538
 QY 596 ctgcgacgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 655
 DB 539 CCCTATTAGAAAGAGCCCAATGACCGTACTATATTTTGGTCTACGCCCAATGCTA 598
 QY 656 gagacgggaagtcggtgttggcaataatctcggactaagcccgagtggtttacacat 715
 DB 599 ATGAAGGAAAATCACAGTTTGGTAAATCTCGGATTAAAGAAATACCTTTATTATTC 658
 QY 716 gtggtgaacacgaagacgatttaccagtcacagcagcagcagcagcagcagcagcagc 775
 DB 659 CTGAGGTAAACCCAGATATGACATATATGTTAATGAAAATCAAGGCAATGTG 718
 QY 776 tcccgatgacacgagtgtaatttagagtatttaattatgcctctgttagaattgttta 835
 DB 719 TGATGGATATCTCTGTTGTAATTCGAATTTAAATTTACCAATTTATGGAATTAATTA 778

QY 836 agaacaggcattcagtcggacaaatcagaaaccccttagttatcttgggttcgaccatg 895
 DB 779 AAAATAGAACCATATATAGTTTAAATATGACACCATGTCGATGATTAATAATAAAA 838
 QY 896 tgcattgactcgtatttgcgaatgcctcctcgtgatttgaataatcagcaggggacagaa 955
 DB 839 TACATGTAATGTTATTAGCTTAATGATTGCTGCAATATGAAAAAATTAGTCAGGATAGAA 898
 QY 956 taaaactgtggaatatttaa 975
 DB 899 TTAATAATAATTATTGTTAA 918
 RESULT 14
 FBEL12187 1007 bp DNA circular VRL 16-SEP-2000
 LOCUS faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian
 DEFINITION EV1-93.
 ACCESSION AJ132187 GI:4995171
 VERSION AJ132187.1
 KEYWORDS C9-Eg gene; rep protein.
 SOURCE faba bean necrotic yellows virus.
 ORGANISM faba bean necrotic yellows virus.
 REFERENCE 1 (bases 1 to 1007)
 AUTHORS Katul, L., Timchenko, T., Gronenborn, B. and Vetter, H.J.
 TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
 JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
 MEDLINE 99094637
 REFERENCE 2 (bases 1 to 1007)
 AUTHORS Timchenko, T., de Kouchkovsky, F., Katul, L., David, C., Vetter, H.J. and Gronenborn, B.
 TITLE A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
 JOURNAL J. Virol. 73 (12), 10173-10182 (1999)
 MEDLINE 20027244
 REFERENCE 3 (bases 1 to 1007)
 AUTHORS Katul, L.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. Land- u. Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
 FEATURES Location/Qualifiers
 1..1007
 /organism="faba bean necrotic yellows virus"
 /viral
 /isolate="Egyptian EV1-93"
 /db_xref="taxon:59817"
 /lab_host="Vicia faba"
 1..39
 73..1003
 /gene="C9-Eg"
 73..918
 /gene="C9-Eg"
 /function="replication-associated protein"
 /note="32.53 kDa"
 /codon_start=1
 /evidence="experimental"
 /product="Rep protein"
 /protein_id="CAB44027.1"
 /db_xref="GI:4995172"
 /db_xref="SPTREMBL:O9WIK2"
 /translation="MSAVNNVFTLNFAAGEVPVLSFDERVOYAVQWHERVNHDIQCVI
 OLKKAKNTVKNITIGGNPHLEKMGKSGIEASAYAKKEESRVAGPWSYGGELKKGSHK
 RKIMELIKDPENLEPEQKYRRAMASAMDESRLAEFGFPYMFYSWQETVLGLLEE
 EPNDRTIIVYGPNGNEKSGQFGKFLGKDKDLYLPQGTQDMYMLMKNPKANVMD
 IPRCNSEYLNQFMELIKNRTIFSYKYPVGCILNNKHIVILANVLPDYKISQDRI
 KLIYC"
 939..944
 /gene="C9-Eg"
 TATA_signal

QY 721 ggaaccagaaggagctattgtaccagttacatcagagaccacacgaacgaatttaactc 780
 Db 657 GGAAGAGCGAGGATATCTGTATATGTATGCTCAAGACCCAGAGAGAAATATCGCGTTC 716
 QY 781 gatgtaccaggtgttaatttagagttatttaattatgcctgttagaattgtgttaagaac 840
 Db 717 GATGTCCAGGTGTCTTCGAGATGATGAACATATCAAGCTATGGAGATGTTGAAGAAC 776
 QY 841 agggcattcagttcggacaataacgaaccccttagttatcttgggttcgaccatgtgcat 900
 Db 777 AGAGTATTTGCAAGTACAAAATATAGGCCCGTAGATCTTTGTGTAGAAAAGTTTCAT 836
 QY 901 gtactcgtatttgcgaatgtcctgcctgattatttgaataatcagcagggacagataaaa 960
 Db 837 TTAATTGTGTTGCAACGTGTACCTGACCCACAAAATAAGTGAGGACAGAAATTGTA 896
 QY 961 ctgtggaatatttaa 976
 Db 897 ATTATCAATTGTTGAA 912

Search completed: October 24, 2001, 11:42:08
 Job time: 6313 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:39 ; Search time 5701.85 Seconds
(without alignments)
2118.665 Million cell updates/sec

Title: US-09-462-955-1_COPY_211_991
Perfect score: 781
Sequence: 1 gttgggacgaggtgcacc.....ttaaagtatgtcatctaa 781

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_in1.*
- 5: gb_in2.*
- 6: gb_in3.*
- 7: gb_om.*
- 8: gb_ov.*
- 9: gb_pat1.*
- 10: gb_pat2.*
- 11: gb_ph.*
- 12: gb_pil.*
- 13: gb_pil2.*
- 14: gb_pil3.*
- 15: gb_pil4.*
- 16: em_ba1.*
- 17: em_ba2.*
- 18: em_fun.*
- 19: em_htgo_hum.*
- 20: em_htgo_inv.*
- 21: em_htgo_rod.*
- 22: em_htg_hum1.*
- 23: em_htg_hum2.*
- 24: em_htg_hum3.*
- 25: em_htg_hum4.*
- 26: em_htg_hum5.*
- 27: em_htg_hum6.*
- 28: em_htg_hum7.*
- 29: em_htg_hum8.*
- 30: em_htg_inv1.*
- 31: em_htg_inv2.*
- 32: em_htg_other.*
- 33: em_htg_rod.*
- 34: em_hum1.*
- 35: em_hum2.*
- 36: em_hum3.*
- 37: em_hum4.*
- 38: em_hum5.*
- 39: em_hum6.*
- 40: em_hum7.*
- 41: em_in.*
- 42: em_om.*
- 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pil.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_vil.*
- 59: gb_vil2.*
- 60: gb_vil3.*
- 61: gb_vil4.*
- 62: gb_vil5.*
- 63: gb_vil6.*
- 64: gb_vil7.*
- 65: gb_vil8.*
- 66: gb_vil9.*
- 67: gb_vil10.*
- 68: gb_vil11.*
- 69: gb_vil12.*
- 70: gb_vil13.*
- 71: gb_vil14.*
- 72: gb_vil15.*
- 73: gb_vil16.*
- 74: gb_vil17.*
- 75: gb_vil18.*
- 76: gb_vil19.*
- 77: gb_vil20.*
- 78: gb_vil21.*
- 79: gb_vil22.*
- 80: gb_vil23.*
- 81: gb_vil24.*
- 82: gb_vil25.*
- 83: gb_vil26.*
- 84: gb_vil27.*
- 85: gb_vil28.*
- 86: gb_vil29.*
- 87: gb_vil30.*
- 88: gb_vil31.*
- 89: gb_vil32.*
- 90: gb_vil33.*
- 91: gb_vil34.*
- 92: gb_vil35.*
- 93: gb_vil36.*
- 94: gb_vil37.*
- 95: gb_vil38.*
- 96: gb_vil39.*
- 97: gb_vil40.*
- 98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	781	100.0	1291	58	CFDCG
2	162.6	20.8	1096	9	AR063452
3	162.6	20.8	1096	58	BYTV2
4	156	20.0	1095	58	AF216222
5	142.6	18.3	1091	9	AR063453
6	142.6	18.3	1111	58	BB012586
7	141	18.1	1127	58	BB012587
8	137.8	17.6	1106	9	AR063451

TRGVKRPRLAORFAEEDELRLDPGGYRRCVVGASVEWTRWAENPFFPPYHMQ
 LEVLSAIGCPADRRILWICRGDGGKSVAKYLGKLPDMFTCTGGTRKDLVQYIE
 DPKRLNLDVPCNLEYLNALLECKNRASSDKYELPSYLGFDHVRHLVFNVLPO
 314..775
 /note="ORF 2"
 /codon_start=1
 /protein_id="AAA42895.1"
 /db_xref="GI:323308"
 /translation="MTGFTWSPRPVPTNRIETVRRNGCFSTESRLVLESKGDHWP
 DLLRNLMNSAKWTQADTEALTELRWQDGPVKIRSHFHTIGSLKCCCLRSQRT
 IAQSSGYADETEETGSPCLPNISDSSTGTHVPEPKTYCTISRTONEI"
 complement(422..568)
 /note="ORF 6"
 /codon_start=1
 /protein_id="AAA42896.1"
 /db_xref="GI:323309"
 /translation="MEMGTFDFORPILSIPPKLRVORIFGIRLGLPGGVHQPQIVGP
 IVAF"
 639..797
 /note="ORF 3"
 /codon_start=1
 /protein_id="AAA42897.1"
 /db_xref="GI:323310"
 /translation="MTRRRRRREVRCVQISRTQARLVLMWNQKGRIVPVHRGP
 FNPRTQV"
 complement(823..987)
 /note="ORF 5"
 /codon_start=1
 /protein_id="AAA42898.1"
 /db_xref="GI:323311"
 /translation="MTHTLNIPQYVSPADPQIIRODICKYEMHMHVEPKITKGFV
 RTECPVLNTE"
 1098..1286
 /note="ORF 4"
 /codon_start=1
 /protein_id="AAA42899.1"
 /db_xref="GI:323312"
 /translation="MNRVMGPGPTIKSIWIRTNLLCLQCTQPLSTSPQVSSLLLEK
 AASLYLPSICFCAIGRLS"
 336 a 323 c 332 g 300 t

L32166 Banana bunc
 AF216221 Banana bu
 AJ005966 faba bean
 AJ132187 faba bean
 U02312 Banana bunc
 AB000921 Milk vetc
 U16731 Subterranea
 AJ132185 faba bean
 AB009047 Milk vetc
 AJ005968 faba bean
 AJ005964 faba bean
 U16735 Subterranea
 X80879 FBIV gene
 AB000922 Milk vetc
 AB000920 Milk vetc
 AJ132345 Nanovirus
 AJ238493 Nanovirus
 AJ132344 Nanovirus
 AR010230 Sequence
 AR112007 Sequence
 AR010227 Sequence
 U18077 Banana bunc
 AR010225 Sequence
 AR010228 Sequence
 AR010240 Sequence
 AF102780 Banana bu
 S56276 orf VI {com
 AR010232 Sequence
 AR010234 Sequence
 AR010229 Sequence
 AR010231 Sequence
 AF238874 Banana bu
 U97525 Banana bunc
 AR010233 Sequence
 AR010226 Sequence
 AF238875 Banana bu
 AF246123 Banana bu

137.8 17.6 1106 58 BYTV1
 136.2 17.4 1109 58 AF216221
 132 16.9 1004 59 NYV5966
 132 16.9 1007 58 FBE132187
 119.6 15.3 1095 59 U02312
 114.6 14.7 1009 58 AB000921
 111.2 14.2 1022 59 SCU16731
 104 13.3 1015 58 FBE132185
 103.4 13.2 1004 59 NYV5968
 101.6 13.0 1014 59 NYV5964
 99.2 12.7 1017 59 SCU16735
 96.4 12.3 1002 58 FBIVREP
 96 12.3 1000 58 AB000922
 92.8 11.9 1007 58 AB000920
 91.4 11.7 1375 58 CLE132345
 88.2 11.3 1367 58 AYE238493
 86.6 11.1 1376 58 CLE132344
 82.8 8.0 1110 9 AR010230
 82.8 7.6 982 9 AR112007
 59.6 7.6 1110 9 AR010227
 59.6 7.6 1110 58 BU18077
 59.6 7.6 1111 9 AR010225
 59.6 7.6 1111 9 AR010228
 59.6 7.6 1111 9 AR010240
 59.6 7.6 1111 58 AF102780
 58.6 7.5 1103 9 S56276
 58.6 7.5 1103 9 AR010232
 58.6 7.5 1105 9 AR010234
 58 7.4 1109 9 AR010229
 58 7.4 1111 9 AR010231
 57 7.3 1103 58 AF238874
 57 7.3 1104 9 U97525
 57 7.3 1103 9 AR010233
 56.4 7.2 1111 9 AR010226
 53.8 6.9 1103 58 AF238875
 52.2 6.7 1103 58 AF246123

RESULT 1
 CFDCG 1291 bp ss-DNA circular VRL
 LOCUS Coconut foliar decay virus, complete genome.
 DEFINITION 02-AUG-1993
 ACCESSION M29963
 VERSION M29963.1 GI:323306
 KEYWORDS circular; complete genome.
 SOURCE Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
 ORGANISM Coconut foliar decay virus
 Viruses; ssDNA viruses; Nanovirus.
 REFERENCE 1 (bases 1 to 1291)
 Rohde,W., Randless,J.W., Langridge,P. and Hanold,D.
 Nucleotide sequence of a circular single-stranded DNA associated
 with coconut foliar decay virus
 JOURNAL Virology 176, 648-651 (1990)
 MEDLINE 90266484
 COMMENT entry and printed sequence for [1] kindly submitted by
 W.Rhode, 15-MAR-1989, for release after publication.
 FEATURES
 source
 /organism="Coconut foliar decay virus"
 /db_xref="taxon:12474"
 stem_loop
 40..70
 /note="stem-loop structure"
 CDS
 103..975
 /note="ORF 1"
 /codon_start=1
 /protein_id="AAA42894.1"
 /db_xref="GI:323307"
 /translation="MGSSIRRWCFILNYETEEAANVRRRIESLNLVAVIGDEVAPS
 TGQRHLQGFILKTRQLGKIVLGNDRHLFETRGSDQNRDYSKERVLLHGVF

ACCESSION U12586
VERSION U12586.1 GI:642390

KEYWORDS Banana bunchy top virus.
SOURCE Banana bunchy top virus
ORGANISM Viruses; ssDNA viruses; Circoviridae.

REFERENCE 1 (bases 1 to 1111)
AUTHORS Wu, R.-Y. and You, L.-R.
TITLE Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related virus DNAs

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1111)
AUTHORS Wu, R.-Y.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1994) Rev-Yuh Wu, Agricultural Biotechnology Division, Development Center for Biotechnology, 81 Chang Hsing Street, Taipei, 10671, Taiwan, ROC

FEATURES
source Location/Qualifiers
1..1111 /organism="Banana bunchy top virus"
/db_xref="taxon:12585"
/clone="H-4"

TATA_signal 1..8
stem_loop 15..49
CDS 79..933
/note="ORF V1"
/codon_start=1
/protein_id="AA61875.1"
/db_xref="GI:642391"
/translation="MSSFKWCFTLNYSSAERDFIALKEELNAYVGVDEVPASS QKHLYSLGKSIKLGLKKKYSSRAHWERARGSDNAKYSKETLLELGFPASQ GNNRLSWVSRSRIPMERIOPELIHYRTYSVKLLKFEEFVPCLDPRWQIQTEA IDEPDRDITFWYVGPNEGEKSTYAKSLMKDWFTRGCKENILFSYVDGESEKHI VFDPICNDYLYNVIEALKDRVIESTKYPKIKLVELNNIHIVMANEMPEFCIKSE DRKIIFYC"
complement(497..643)
/codon_start=1
/product="ORF C1"
/protein_id="AA61876.1"
/db_xref="GI:642392"
/translation="MTSHMLISPHVHSDRPFRSFHDHLVPRQLPPSESAMVVRGKDE RIP"

BASE COUNT 344 a 213 c 260 g 294 t
ORIGIN

Query Match	18.3%;	Score 142.6;	DB 58;	Length 1111;
Best Local Similarity	52.99;	Pred. No. 2.5e-31;		
Matches 407;	Conservative	0;	Mismatches 344;	Indels 18; Gaps 4;

```

Qy      1  gtggggagcagggttgccacctcaactcgccgaacgacacacctccaaggattccacctg   60
          ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     181  GTCGGGCGAGAGTGCCTCCGAGTCGCGGCCGAAGAAGCACCTACAGGGATATCATCCCTG   240
          ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      61  aaagaccgcgcgcagtcaaggattgaagactgtcttctgggaatcacaggattccacctg   120
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Db     241  AAGAAATCTATAAACCTTGAGGATTGAAGAAGAAGTAT--TCTTCGAGAGTCACTGG    297
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Qy     121  gagccaccctgtgttcgcgaacagaataagactactgttcgaagaacgggtgctt       180
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Db     298  GAGAGGGCGGAGAGGATCTGACCAGATTAATGCTTAAGTACTGTTCCTTAAGAACCCCTAATT   357
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Qy     181  ctcgagcacggagtcgccgactgcctggactgcagaaagccagcagattggcccacaagatt   240
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Db     358  CTGGAATTGGGGTTTCCCCTCTCAAGGCTCTAATAGACGGAACAATATCGGAGATGGTT    417
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Qy     241  gctgaggaaactgatgaactccgcctctggaagaccaccgagcgagataccgaagaatcgttga   300
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Db     418  TCTAGATCTCCAGAACGCATGAGATTGAACAGCGCTGAGATATATCACAGAT-----    469
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||
Qy     301  cacggaacttcggtggaatggacaagatgggcgcgtgaaaaatccgttcccatcttcacat   360
          ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||| |||

```

470	-ACACATCTGTGAAGAAGTTAAAAAATAATCAAGGAGGAATTCGTTTCATCTTCGCCTCGAT	528
361	cacaattggcagctgaagtgcctctgcgatcgagaccagcgacgatccacaatc	420
529	AGACCATGGCAGATTCAAATTCACGAGGCAATTCACGAGGACACAGATGATCGAAGATC	588
421	ctctcgatcgcgacgagacgaggaagtcggtttgttgccaaatatctcgga	480
589	TTCCTGGTCTATGTCGCAATGGTAATCAGGGGAAATCAACATATCGAAGTCATT---	645
481	ctcagcgcactggtctcacatggtgtgaaccagaagcgctatgtaccagtac	540
646	ATCAAGAGGACTGTGTTTACACAGAGGTGGAGAGGAGACATACTGTCTCTTAC	705
541	atcagaggaccccaaacgaa---atttaactctcgatgtaccacgggtgaatttagagtat	597
706	GTGACGAGGATCTGAGAAGCATATTCATTGATATTCCTCGCTTAATCAAGATTAT	765
598	ttaattatgcctgtttagaatgtttaagaacagggcattcagttcggacaatacga	657
766	TTAAATATATGATTATAGAGGCATTAAGAGGATAGGGTCATAGAGAGTACTAATATAAA	825
658	cccttagtattcttggttcgaccatgtgcactctactcgtattgccaatgtcctgcct	717
826	CCTATTAACTTAGTTCGAATGAATAATACATGTAATTGTAATGGCTAATTTCATGCCA	885
718	gattattgaaatacgcgaggaagacagaataaacgtcggaattatttaa	766
886	GAATTCTGTAAATCTCCGAAGATAGATAAGATATTTATTGTGTAA	934
RESULT 7		
BBU12587	BBU12587 1127 bp DNA circular VRL 01-FEB-1995	
LOCUS	Banana bunchy top virus DNA IV ORF VI and ORF C1 genes, complete cds.	
ACCESSION	U12587	
VERSION	U12587.1 GI:642393	
KEYWORDS	Banana bunchy top virus.	
SOURCE	Banana bunchy top virus.	
ORGANISM	Banana bunchy top virus.	
REFERENCE	Viruses; ssDNA viruses; Circoviridae.	
AUTHORS	Wu, R.-Y. and You, L.-R.	
TITLE	Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related viruses	
JOURNAL	DNAS	
REFERENCE	Unpublished	
AUTHORS	Wu, R.-Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (25-JUL-1994) Rey-Yuh Wu, Agricultural Biotechnology Division, Development Center for Biotechnology, 81 Chang Hsing Street, Taipei, 10671, Taiwan, ROC	
FEATURES	Location/Qualifiers	
source	1..1127	
	/organism="Banana bunchy top virus"	
	/db_xref="taxon:12585"	
	/clones="SI5 and 4-2"	
TATA_signal	1..8	
stem_loop	15..49	
CDS	79..933	
	/note="ORF V1"	
	/codon_start=1	
	/protein_id="AAA61877.1"	
	/db_xref="GI:642394"	
	/translation="MSSFKWCFTLNTYSSAEREDFALLKEBELNAVVGDEVAPSSG QKHLOGYLSLKKSTIKGLKKYSSRAHWRARGSDENAKYCSKFTLLLELGFPAQ GSNRKLSEWVSRSPEWRIQEPFIYHRVTSVKLLKFKDEFVHPCLDPRWQIQLETA IDEPDRQDILFWITGPNNGNECKSTYAKSLMKDWFTRGCKENILFYSVDGSKH1 VFDTPRCNQYLV1EALCKADRVISTKTKPKIKLVNLNIHVIVMANFMPRECKISE DRIKIYC"	
	complement(497..643)	
CDS		

Thu Oct 25 13:07:56 2001

```

RESULT 9
BYTV1 1106 bp DNA VRL 30-OCT-1994
LOCUS Banana bunchy top virus (BEV DNA I) V1 and C1-C3 genes, complete
DEFINITION cds's.
ACCESSION L32166
VERSION L32166.1 GI:520786
KEYWORDS stem loop.
SOURCE Banana bunchy top virus DNA.
ORGANISM Bananas; ssDNA viruses; Circoviridae.
REFERENCE Wu, R.-Y., You, L.-R. and Soong, T.-S.
AUTHORS Nucleotide sequences of two circular single-stranded DNAs
TITLE associated with banana bunchy top virus
JOURNAL Unpublished (1994)
FEATURES
    Location/Qualifiers
        1..1106
            /organism="Banana bunchy top virus"
            /db_xref="taxon:12585"
        1..7
            /TATA_signal
        9..37
            /stem_loop
        62..922
            /gene="V1"
        62..922
            /gene="V1"
        /codon_start=1
        /protein_id="AAAS1422.1"
        /db_xref="GI:520787"
        /translation="MSPFSLKWCFTLNYSAAERENFLSKKEEDVHYAVVGEVAPA
        TQKHLOGLSLKRIKRLGKKYKYSRAHWEIARGTDENSKYCKSKETLLELGPV
        VNGNKRKISEMVARSPDKIEQPEIFHRYQSVNKLKPKFEFVHPCLDRWQIQLT
        PAIDEEPDRSIWVYGPYNEGKSTYAKSLIKKDWFTYRGGKKNILFSYVDEGSDK
        HIVPDIKPCNDYLVNDVIEALKDRVIESTKYKPKIKIVELGKHIVMNFMPDFCKI
        SEDRKIIYC"
        complement(280..426)
        /gene="C3"
        complement(280..426)
        /gene="C3"
        /codon_start=1
        /protein_id="AAAS1423.1"
        /db_xref="GI:520788"
        /translation="MRSERATISDFILLEPLTGNPNRIRVSEFQYFERSSSVPL
        AISQ"
        complement(636..779)
        /gene="C1"
        complement(636..779)
        /gene="C1"
        /codon_start=1
        /protein_id="AAAS1424.1"
        /db_xref="GI:520789"
        /translation="MPLHLNLLNPDYNEEVOILYACQILRPHRRIRYSPSSHPWCT
        NPS"
        complement(760..912)
        /gene="C2"
        complement(760..912)
        /gene="C2"
        /codon_start=1
        /protein_id="AAAS1425.1"
        /db_xref="GI:520790"
        /translation="MIFIRSRSEILQSGMKFAMTTCILPNSTIFMGLYLVLSTLSF
        NASINS"
BASE COUNT 335 a 225 c 257 g 289 t
ORIGIN

```

```

Query Match 17.6%; Score 137.8; DB 58; Length 1106;
Best Local Similarity 50.7%; Pred. No. 6.6e-30;
Matches 390; Conservative 0; Mismatches 367; Indels 12; Gaps 2;
Qy 1 gttgggacgaggttgacattcaactgccaacgacacacacgaggtccacttg 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 170 GTCGCGACGGAAGTCGCTCCGGCCACCGGCCAGAAAGACACCTCCAGGGATATCTATCCCTG 229
Qy 61 aagaccggttcggcgactgcaagattgaagactgttcttgggaatgacaggttacctg 120
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 230 AAAAAGAGATCCGCTCGCGGATTTGAAGAAGAAAGATATGG---TTCCCGTGCACATGG 286
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 gaccgaccggtgttccgacgaacaagaatagagactactgttcgaaggaacggtggtt 180
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 287 GAGATCCAGAGACGAGACGAGAGAAATTCGAAGTACTGTTCAAAAAGAAACCCCTAAT 346
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 ctgagcagcaggtcccgactgctcctggagtcgaagaagccagcttggcccaacgatt 240
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 CTCGAATTAGGGTTTCCTGTTTAAATGTTCTTAATAAAGGAAAAATATCGGAGATGGT 406
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 gctgagaaactgatgaactcccgctggaagaccgagcggaataccgaagatgcttgta 300
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 407 GCTGTTCTCCTGATCGCATGAAAAATTGAACGCCCTGAGATATTTTCACAGAT----- 458
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 caccgagcttcggtgggaatggacaagaatggcgctgaaataccgttcccatattccat 360
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 459 -ATCAATCTGTAATAAGTTAAAAAATTTCAAGGAGAGTTTCCTTCATCCTTGCCTCGAT 517
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 361 cacaattggcagcttggaagtgcgtcgtcgatcgagagccgagcagcagcagcagcagc 420
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 AGACCATGGCAGATTCATTTGACGGAGGCAATTGACGAGGAACCCGATGATCGAAGCATC 577
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 ctctgatatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 ATCTGGGTCTATGTCCTTATGTTAATGAGGTAAATCAACATATGCGAAGTCACTAATC 637
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 481 ctcaagcccgactggttctacacatgtgttggaaccgagcagcagcagcagcagcagcagc 540
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 AAGAAGGATTTGGTCTTACACCAGGGGTGGGAAGAGAGAAATATCTTATTTCTCCTATGTG 697
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 541 atcaggagcccaaacgaatttaactcgtgatacccgagtgtaatttagagatttta 600
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 698 GACGAAGATCTGACACGATATAGTATTGATGATTTCTGTTGTAATCAGGATTTATTA 757
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 601 aattatgcccctgttagaattgttlaagaacagggcattcagttcggacaataacgaacc 660
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 758 AATTATGATGTAATAGAGGCATTTAAAGGATAGGCTTATAGAGAGTAAATACAAACCC 817
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 661 cttagttatcttggttgcagacatgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 720
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 818 ATAAAGATAGTTGAATTAGTAAATACATGTAATCGTCATGCGCAATTTTCATGCTTGAC 877
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 721 tattgaaaaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 769
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 878 TTCTGTAAATCTCCGAGATCGAATAAATCAATCATTTATTGCTGAAGAA 926
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

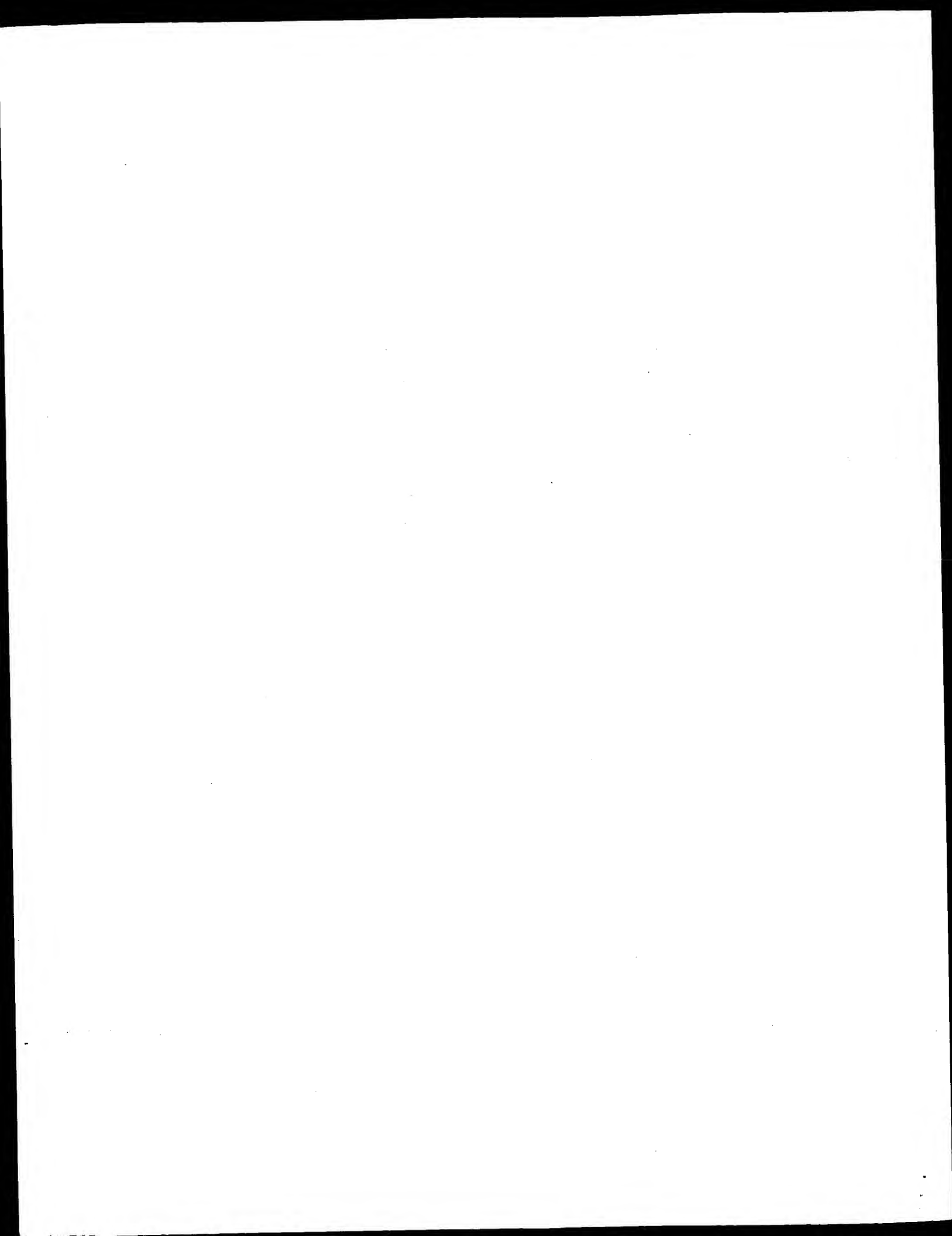
```

RESULT 10
AF216221 1109 bp DNA VRL 08-MAR-2001
LOCUS Banana bunchy top virus satellite s1 replication initiation protein
DEFINITION (ORF V1) gene, complete cds.
ACCESSION AF216221
VERSION AF216221.1 GI:12004325
KEYWORDS Bananas; ssDNA viruses; Circoviridae.
SOURCE Banana bunchy top virus.
ORGANISM Bananas; ssDNA viruses; Circoviridae.
REFERENCE 1 (bases 1 to 1109)
AUTHORS Horser, C., Harding, R. and Dale, J.
TITLE Bananas bunchy top nanovirus DNA-1 encodes the 'master' replication
INITIATION protein
JOURNAL J. Gen. Virol. 82 (Pt 2), 459-464 (2001)
MEDLINE 21102983
PUBMED 11161286
REFERENCE 2 (bases 1 to 1109)
AUTHORS Horser, C.L., Karan, M., Harding, R.M. and Dale, J.L.
TITLE Additional rep-encoding DNAs associated with banana bunchy top
virus

```


[illegible]

RESULT 15
SSCU16731
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:58 ; Search time 5701.85 Seconds
(without alignments)
24.415 Million cell updates/sec

Title: US-09-462-955-5
Perfect score: 9
Sequence: 1 ctagtattta 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl :

1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_ro.*
50: em_ro.*
51: em_sy.*
52: em_un.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_v11.*
59: gb_v12.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_ro1.*
95: gb_ro2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match %	Length	DB ID	Description
C	1	9	100.0	12	9 AR063463	AR063463 Sequence
	2	9	100.0	22	10 AX069023	AX069023 Sequence
	3	9	100.0	30	9 AR097505	AR097505 Sequence
	4	9	100.0	32	9 A93671	A93671 Sequence 7
	5	9	100.0	35	9 A62983	A62983 Sequence 2
C	6	9	100.0	36	10 E22884	E22884 DNA sequence
	7	9	100.0	38	9 AR093937	AR093937 Sequence
	8	9	100.0	38	10 AX081631	AX081631 Sequence

Thu Oct 25 13:08:56 2001

us-09-462-955-5.rge

```

c 9 100.0 38 10 E22871 DNA sequence
10 9 100.0 39 10 AX081623
11 9 100.0 41 9 AR104937 Sequence
12 9 100.0 45 9 A39423
13 9 100.0 62 9 A39327
14 9 100.0 74 10 AX069021
15 9 100.0 81 58 AF227763
16 9 100.0 86 97 HSZ99564
17 9 100.0 94 2 C20133579
18 9 100.0 94 2 C20133579
19 9 100.0 101 10 I12330
20 9 100.0 101 10 I70233
21 9 100.0 103 88 AF055272
22 9 100.0 106 54 G32785
23 9 100.0 111 2 C20133577
24 9 100.0 111 59 HIV1M00951
25 9 100.0 114 97 HUMPERNAC
26 9 100.0 117 54 G00157
27 9 100.0 120 14 S76763
28 9 100.0 121 58 AF203205
29 9 100.0 121 58 AF203214
30 9 100.0 121 93 HSPA16F10
31 9 100.0 123 3 RBKAN2
32 9 100.0 126 3 SSISSAMB
33 9 100.0 126 93 HSAX1532
34 9 100.0 127 3 PBOKAN2
35 9 100.0 127 58 AF203217
36 9 100.0 129 3 INS2MUTA
37 9 100.0 130 3 INS43AAA
38 9 100.0 133 58 AF203212
39 9 100.0 136 54 G44046
40 9 100.0 138 59 HIVU76695
41 9 100.0 139 54 G34972
42 9 100.0 141 5 AF068581
43 9 100.0 141 5 HSPE05H04
44 9 100.0 142 3 LACSS38AA
45 9 100.0 143 1 AB005263

```

ALIGNMENTS

```

RESULT 1
LOCUS AR063463 12 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 15 from patent US 5846705.
ACCESSION AR063463
VERSION AR063463.1 GI:5992771
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12)
AUTHORS Wu, R., You, L. and Soong, T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus
JOURNAL Patent: US 5846705-A 15 08-DEC-1998;
FEATURES Location/Qualifiers
source 1..12
BASE COUNT 3 a 4 c 1 g 4 t
ORIGIN

```

```

Query Match 100.0%; Score 9; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
Db 2 CTAGTATTATA 10

```

```

RESULT 2
LOCUS AX069023/c 22 bp DNA PAT 25-JAN-2001
DEFINITION Sequence 3 from Patent WO0102555.
ACCESSION AX069023
VERSION AX069023.1 GI:12578838
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
REFERENCE 1 (bases 1 to 22)
AUTHORS Gicquel, B., Guillehot, C. and Canacho, L.
TITLE Method of making and identifying attenuated microorganisms, compositions utilizing the sequences responsible for attenuation, and preparations containing attenuated microorganisms
JOURNAL Patent: WO 0102555-A 3 11-JAN-2001;
FEATURES Location/Qualifiers
source 1..22
/organism="Mycobacterium tuberculosis"
BASE COUNT 7 a 5 c 1 g 9 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.9e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
Db 11 CTAGTATTATA 3

RESULT 3
LOCUS AR097505 30 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 33 from patent US 6071732.
ACCESSION AR097505
VERSION AR097505.1 GI:12806235
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Moore, K.L.
TITLE Tyrosylprotein sulfotransferases, nucleic acids encoding tyrosylprotein sulfotransferases, and methods of use thereof
JOURNAL Patent: US 6071732-A 33 06-JUN-2000;
FEATURES Location/Qualifiers
source 1..30
/organism="unknown"
BASE COUNT 5 a 9 c 6 g 10 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 30;
Best Local Similarity 100.0%; Pred. No. 9.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
Db 4 CTAGTATTATA 12

RESULT 4
LOCUS A93671 32 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9734144.
ACCESSION A93671

```

```

VERSION      A93671.1  GI:6741859
SOURCE       synthetic construct.
ORGANISM     synthetic construct
REFERENCE    1 (bases 1 to 32)
AUTHORS     Langer, G. and Yoshida, L.
TITLE       METHOD OF DETECTING THE EFFECT OF TEST SUBSTANCES USING HEN
JOURNAL     UROKINASE
PATENT      WO 9734144-A 7 18-SEP-1997;
SCHEDING AG (DE): LANGER GERNOT (DE)
FEATURES     Location/Qualifiers
              source
                1..32
                /organism="synthetic construct"
                /db_xref="taxon:32630"
BASE COUNT   5 a      7 c      11 g      9 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
262
```

Thu Oct 25 13:08:56 2001

us-09-462-955-5.ige

Db 18 CTAGTATTA 10
 RESULT 10
 AX081623 39 bp DNA PAT 27-FEB-2001
 LOCUS Sequence 128 from Patent WO0109350.
 DEFINITION AX081623
 ACCESSION AX081623.1 GI:13170448
 VERSION
 KEYWORDS synthetic construct.
 ORGANISM synthetic construct
 SOURCE artificial sequence.
 REFERENCE 1 (bases 1 to 39)
 AUTHORS Lobet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
 Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
 TITLE Genetically engineered bleb vaccine
 JOURNAL Patent: WO 0109350-A 128 08-FEB-2001;
 SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
 FEATURES Location/Qualifiers
 source 1..39
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="BAD 03-2 primer"
 BASE COUNT 6 a 14 c 10 g 9 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 10; Length 39;
 Best Local Similarity 100.0%; Pred. No. 8.7e+04; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ctagtattta 9
 Db 18 CTAGTATTA 27
 RESULT 11
 AX081623 41 bp DNA PAT 14-FEB-2001
 LOCUS Sequence 106 from patent US 6096321.
 DEFINITION AX081623
 ACCESSION AX081623.1 GI:12818534
 VERSION
 KEYWORDS synthetic construct.
 ORGANISM synthetic construct
 SOURCE artificial sequence.
 REFERENCE 1 (bases 1 to 41)
 AUTHORS Lobet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
 Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
 TITLE Genetically engineered bleb vaccine
 JOURNAL Patent: WO 0109350-A 128 08-FEB-2001;
 SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
 FEATURES Location/Qualifiers
 source 1..41
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="BAD 03-2 primer"
 BASE COUNT 11 a 3 c 10 g 17 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 9; Length 41;
 Best Local Similarity 100.0%; Pred. No. 8.6e+04; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ctagtattta 9
 Db 1 CTAGTATTA 9
 RESULT 12
 A39423 45 bp DNA PAT 05-MAR-1997
 LOCUS Sequence 42 from Patent WO9414967.
 DEFINITION

Db 18 CTAGTATTA 10
 RESULT 10
 AX081631 38 bp DNA PAT 07-FEB-2001
 LOCUS DNA sequence increasing promoter activity.
 DEFINITION AX081631
 ACCESSION AX081631.1 GI:13024149
 VERSION
 KEYWORDS JP 1999075843-A/8.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 38)
 AUTHORS Toshihiro,Y.
 TITLE DNA sequence increasing promoter activity
 JOURNAL Patent: JP 1999075843-A 8 23-MAR-1999;
 COMMENT KIRIN BREWERY CO LTD
 OS Unidentified
 PN JP 1999075843-A/8
 PD 23-MAR-1999
 PF 29-AUG-1997 JP 1997234995
 PR TOSHIHIRO,YONEDA
 PC C12N15/09,C12N1/19,C12P21/02,((C12N15/09,C12R1:72),(C12N1/19,
 PC C12R1:72),
 PC C12P21/02,C12R1:72),C12N15/00,(C12N15/00,C12R1:72) CC
 Strandedness: Single;
 CC Topology: Linear;
 FH Key Location/Qualifiers
 FT source 1..38
 /organism="Unidentified".
 FEATURES Location/Qualifiers
 source 1..38
 /organism="Unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 12 a 8 c 5 g 13 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 10; Length 38;
 Best Local Similarity 100.0%; Pred. No. 8.8e+04; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ctagtattta 9
 Db 18 CTAGTATTA 26

ACCESSION A39423
 VERSION A39423.1 GI:2295756
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 45)
 AUTHORS Girardeau, J., Martin, C., Mechin, M., Der, V. M. and Bousquet, F.
 TITLE CS31A PROTEIN CAPSULE SUBUNIT MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A PROTEIN CAPSULE COMPRISING SAID SUBUNIT, MICRO-ORGANISMS HAVING SUCH SUBUNITS ON THEIR OUTER MEMBRANE, AND METHODS FOR PREPARING AND USING SAME
 JOURNAL Patent: WO 9414967-A 42 07-JUL-1994;
 COMMENT AGRONOMIQUE INST NAT RECH (FR)
 Other publication AU 6808794 940719
 Other publication FR 2699538 940624.
 FEATURES
 source 1..45
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 12 a 4 c 11 g 18 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 9; Length 45;
 Best Local Similarity 100.0%; Pred. No. 8.5e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ctagtattta 9
 |||||
 Db 1 CTAGTATTTA 9
 RESULT 13
 LOCUS A39327 62 bp DNA PAT 05-MAR-1997
 DEFINITION Sequence 9 from Patent WO9414969.
 ACCESSION A39327
 VERSION A39327.1 GI:2295680
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 62)
 AUTHORS Adams, S.E., Burns, R.N. and Richardson, S.M.
 TITLE NOVEL PROTEINACEOUS PARTICLES
 JOURNAL Patent: WO 9414969-A 9 07-JUL-1994;
 BRITISH BIO TECHNOLOGY (GB)
 COMMENT Other publication AU 5711254 940719
 Other publication JP 8505788T 960625.
 FEATURES
 source 1..62
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 23 a 12 c 10 g 17 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 9; Length 62;
 Best Local Similarity 100.0%; Pred. No. 7.9e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ctagtattta 9
 |||||
 Db 1 CTAGTATTTA 9
 RESULT 14
 LOCUS AX069021 74 bp DNA PAT 25-JAN-2001
 DEFINITION Sequence 1 from Patent WO0102555.
 ACCESSION AX069021
 VERSION AX069021.1 GI:12578836

KEYWORDS
 SOURCE Mycobacterium tuberculosis.
 ORGANISM Mycobacterium tuberculosis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 REFERENCE 1 (bases 1 to 74)
 AUTHORS Gicquel, B., Guilhot, C. and Camacho, L.
 TITLE Method of making and identifying attenuated microorganisms, compositions utilizing the sequences responsible for attenuation, and preparations containing attenuated microorganisms
 JOURNAL Patent: WO 0102555-A 1 11-JAN-2001;
 INSTITUT PASTEUR (FR)
 FEATURES
 source 1..74
 /organism="Mycobacterium tuberculosis"
 /db_xref="taxon:1773"
 BASE COUNT 21 a 9 c 8 g 16 t 20 others
 ORIGIN
 Query Match 100.0%; Score 9; DB 10; Length 74;
 Best Local Similarity 100.0%; Pred. No. 7.6e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ctagtattta 9
 |||||
 Db 64 CTAGTATTTA 72
 RESULT 15
 LOCUS AF227763 81 bp RNA VRL 12-JAN-2001
 DEFINITION Hepatitis C virus isolate C10 envelope protein 2 (E2) gene, partial cds.
 ACCESSION AF227763
 VERSION AF227763.1 GI:8099446
 KEYWORDS
 SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus.
 REFERENCE 1 (bases 1 to 81)
 AUTHORS Ross, R.S., Viazov, S., Gross, T., Hofmann, F., Seipp, H.M. and Roggendorf, M.
 TITLE Transmission of the hepatitis C virus from a patient to an anesthesiology assistant to five patients
 JOURNAL N. Engl. J. Med. 343 (25), 1851-1854 (2000)
 MEDLINE 21004055
 PUBMED 11117977
 REFERENCE 2 (bases 1 to 81)
 AUTHORS Ross, S., Viazov, S. and Roggendorf, M.
 TITLE Direct Submission
 JOURNAL Submitted (21-JAN-2000) Institute of Virology, Essen University Hospital, Hufelandstr. 55, Essen D-45122, Germany
 FEATURES
 source 1..81
 /organism="Hepatitis C virus"
 /isolate="C10"
 /db_xref="taxon:11103"
 /note="genotype: 1a"
 <1..>81
 /gene="E2"
 <1..>81
 /gene="E2"
 /note="hypervariable region 1; HVR1"
 /codon_start=1
 /product="envelope protein 2"
 /protein_id="AA072123.1"
 /db_xref="GI:8099447"
 /translation="HPTVTGGSAAYTSSITRLTPGAKON"
 BASE COUNT 21 a 29 c 18 g 13 t
 ORIGIN

Thu Oct 25 13:08:56 2001

Query Match 100.0%; Score 9; DB 58; Length 81;
Best Local Similarity 100.0%; Pred. No. 7.5e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
|||||
Db 41 CTAGTATTTA 49

Search completed: October 24, 2001, 11:42:58
Job time: 6363 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:55 ; Search time 5701.85 Seconds
(without alignments)
24.415 Million cell updates/sec

Title: US-09-462-955-4
Perfect score: 9
Sequence: 1 taactactag 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_bal:*

2: gb_bal:*

3: gb_bal:*

4: gb_bal:*

5: gb_bal:*

6: gb_bal:*

7: gb_bal:*

8: gb_bal:*

9: gb_bal:*

10: gb_bal:*

11: gb_bal:*

12: gb_bal:*

13: gb_bal:*

14: gb_bal:*

15: gb_bal:*

16: gb_bal:*

17: gb_bal:*

18: gb_bal:*

19: gb_bal:*

20: gb_bal:*

21: gb_bal:*

22: gb_bal:*

23: gb_bal:*

24: gb_bal:*

25: gb_bal:*

26: gb_bal:*

27: gb_bal:*

28: gb_bal:*

29: gb_bal:*

30: gb_bal:*

31: gb_bal:*

32: gb_bal:*

33: gb_bal:*

34: gb_bal:*

35: gb_bal:*

36: gb_bal:*

37: gb_bal:*

38: gb_bal:*

39: gb_bal:*

40: gb_bal:*

41: gb_bal:*

42: gb_bal:*

43: gb_bal:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vl:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vl:*

59: gb_vl2:*

60: gb_hgt1:*

61: gb_hgt2:*

62: gb_hgt3:*

63: gb_hgt4:*

64: gb_hgt5:*

65: gb_hgt6:*

66: gb_hgt7:*

67: gb_hgt8:*

68: gb_hgt9:*

69: gb_hgt10:*

70: gb_hgt11:*

71: gb_hgt12:*

72: gb_hgt13:*

73: gb_hgt14:*

74: gb_hgt15:*

75: gb_hgt16:*

76: gb_hgt17:*

77: gb_hgt18:*

78: gb_hgt19:*

79: gb_hgt20:*

80: gb_hgt21:*

81: gb_hgt22:*

82: gb_hgt23:*

83: gb_hgt24:*

84: gb_hgt25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_ro1:*

95: gb_ro2:*

96: gb_in4:*

97: gb_pr10:*

98: em_bal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	9	100.0	12	9	AR063463
C 2	9	100.0	22	10	AX069023
C 3	9	100.0	30	9	AR097505
C 4	9	100.0	32	9	A93671
C 5	9	100.0	35	9	A62983
C 6	9	100.0	36	10	E22884
C 7	9	100.0	38	9	AR093937
C 8	9	100.0	38	10	AX081631

9 100.0 38 10 E22871 DNA sequenc
c 10 9 100.0 39 10 AX081623 Sequence
c 11 9 100.0 41 9 AR104937
c 12 9 100.0 45 9 A39423
c 13 9 100.0 62 9 A39327
c 14 9 100.0 74 10 AX069021
c 15 9 100.0 81 58 AF227763 Hepatitis
c 16 9 100.0 86 97 HS299564
c 17 9 100.0 94 2 CC0133579
c 18 9 100.0 94 2 CJF133576
c 19 9 100.0 101 10 I12330
c 20 9 100.0 101 10 I70233 Sequence 14
c 21 9 100.0 103 88 AF055272 Homo sapi
c 22 9 100.0 106 54 G32785
c 23 9 100.0 111 2 CJF133577
c 24 9 100.0 111 59 HIV1M009S1
c 25 9 100.0 114 97 HUMPSRNAC
c 26 9 100.0 117 54 G00157
c 27 9 100.0 120 14 S76763
c 28 9 100.0 121 58 AF203205
c 29 9 100.0 121 58 AF203214
c 30 9 100.0 121 93 HSPAL6F10
c 31 9 100.0 123 3 RBKXAN2
c 32 9 100.0 126 3 SSISSAMB
c 33 9 100.0 126 93 HSAOX1S32
c 34 9 100.0 127 3 PR0KXAN2
c 35 9 100.0 127 58 AF203217
c 36 9 100.0 129 3 INS2MUTA
c 37 9 100.0 130 3 INS43AAA
c 38 9 100.0 133 58 AF203212
c 39 9 100.0 136 54 G44046
c 40 9 100.0 138 59 HIVU76695
c 41 9 100.0 139 54 G34372
c 42 9 100.0 141 5 AF068581
c 43 9 100.0 141 54 HSP05H04
c 44 9 100.0 142 3 LACS38AA
c 45 9 100.0 143 1 AB005263

ALIGNMENTS

RESULT 1
LOCUS AR063463/c 12 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 15 from patent US 5846705.
ACCESSION AR063463
VERSION AR063463.1 GI:5992771
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 12)
AUTHORS Wu, R., You, L. and Soong, T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus
JOURNAL Patent: US 5846705-A 15 08-DEC-1998;
FEATURES Location/Qualifiers
source 1..12
BASE COUNT 3 a 4 c 1 g 4 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
Db 10 TAATACTAG 2

RESULT 2
LOCUS AX069023 22 bp DNA PAT 25-JAN-2001
DEFINITION Sequence 3 from Patent WO0102555.
ACCESSION AX069023
VERSION AX069023.1 GI:12578838
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
REFERENCE 1 (bases 1 to 22)
AUTHORS Gicquel, B., Guilhot, C. and Camacho, L.
TITLE Method of making and identifying attenuated microorganisms, compositions utilizing the sequences responsible for attenuation, and preparations containing attenuated microorganisms
JOURNAL Patent: WO 0102555-A 3 11-JAN-2001;
INSTITUT PASTEUR (FR)
FEATURES Location/Qualifiers
source 1..22
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
BASE COUNT 7 a 5 c 1 g 9 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.9e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
Db 3 TAATACTAG 11

RESULT 3
LOCUS AR097505/c 30 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 33 from patent US 6071732.
ACCESSION AR097505
VERSION AR097505.1 GI:12806235
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 30)
AUTHORS Moore, K.L.
TITLE Tyrosylprotein sulfotransferases, nucleic acids encoding tyrosylprotein sulfotransferases, and methods of use thereof
JOURNAL Patent: US 6071732-A 33 06-JUN-2000;
FEATURES Location/Qualifiers
source 1..30
/organism="unknown"
BASE COUNT 5 a 9 c 6 g 10 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 30;
Best Local Similarity 100.0%; Pred. No. 9.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
Db 12 TAATACTAG 4

RESULT 4
LOCUS A93671/c 32 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 7 from Patent WO9734144.
ACCESSION A93671

```

VERSION      A93671.1  GI:6741859
KEYWORDS
SOURCE       synthetic construct.
ORGANISM     synthetic construct
              artificial sequence.
REFERENCE    1 (bases 1 to 32)
AUTHORS     Langer,G. and Toschi,L.
TITLE       METHOD OF DETECTING THE EFFECT OF TEST SUBSTANCES USING HEN
              UROKINASE
JOURNAL
FEATURES
SOURCE      Patent: WO 9734144-A 7 18-SEP-1997;
              SCHERING AG (DE); LANGER GERNOT (DE)
              Location/Qualifiers
                1..32
                /organism="synthetic construct"
                /db_xref="taxon:32630" 9 t
BASE COUNT   5 a 7 c 11 g
ORIGIN
Query Match 100.0%; Score 9; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 9.1e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 20 TAATACTAG 12

RESULT 5
LOCUS      A62983
DEFINITION Sequence 2 from Patent WO9720056.
ACCESSION A62983
VERSION   A62983.1  GI:3716855
KEYWORDS  pea.
SOURCE    Pisum sativum
          Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
          Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
          Fabales; Fabaceae; Papilionoideae; Pisum.
REFERENCE  1 (bases 1 to 35)
AUTHORS   Gray,J.C., Sandhu,J.S. and Webster,C.I.
TITLE     ENHANCED GENE EXPRESSION IN PLANTS
JOURNAL   Patent: WO 9720056-A 2 05-JUN-1997;
          CAMBRIDGE ADVANCED TECH (GB)
COMMENT   Other publication AU 7635696 19970619.
FEATURES
SOURCE    1..35
          /organism="Pisum sativum"
          /db_xref="taxon:3888"
          /clone_lib="GENOMIC, LAMEDA EMBL3"
          /clone_lib="(X)SEQUENCE DESCRIPTION:SEQ ID NO: 2:"
BASE COUNT 17 a 4 c 2 g 12 t
ORIGIN
Query Match 100.0%; Score 9; DB 9; Length 35;
Best Local Similarity 100.0%; Pred. No. 8.9e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 19 TAATACTAG 11

RESULT 6
LOCUS      E22884
DEFINITION DNA sequence increasing promoter activity.
ACCESSION E22884
VERSION   E22884.1  GI:13024162
KEYWORDS  JP 1999075843-A/21.
SOURCE    unidentified.

VERSION      A93671.1  GI:6741859
KEYWORDS
SOURCE       synthetic construct.
ORGANISM     synthetic construct
              artificial sequence.
REFERENCE    1 (bases 1 to 36)
AUTHORS     Toshihiro, Y.
TITLE       DNA sequence increasing promoter activity
JOURNAL     Patent: JP 1999075843-A 21 23-MAR-1999;
              KIRIN BREWERY CO LTD
COMMENT     OS Unidentified
              PN JP 1999075843-A/21
              PD 23-MAR-1999
              PE 29-AUG-1997 JP 1997234995
              PR
              PI TOSHIHIRO YONEDA
              PC C12N15/09,C12N1/19,C12P21/02//((C12N15/09,C12R1:72),(C12N1/19,
              PC C12R1:72)),
              PC (C12P21/02,C12R1:72),C12N15/00,(C12N15/00,C12R1:72) CC
              Strandedness: Single;
              CC Topology: Linear;
              FH Key Location/Qualifiers
              FT source 1..36
              FT /organism="Unidentified".
FEATURES
SOURCE    1..36
          Location/Qualifiers
            1..36
            /organism="unidentified"
            /db_xref="taxon:32644" 12 t
BASE COUNT 13 a 8 c 3 g
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 36;
Best Local Similarity 100.0%; Pred. No. 8.9e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 34 TAATACTAG 26

RESULT 7
LOCUS      AR093937
DEFINITION Sequence 41 from patent US 6001590.
ACCESSION AR093937
VERSION   AR093937.1  GI:10020682
KEYWORDS  Unknown.
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE  1 (bases 1 to 38)
AUTHORS   Komeda,T., Suda,H., Tamai,Y., Iwamatsu,A., Kato,N. and Sakai,Y.
TITLE     Promoter and terminator sequences of formate dehydrogenase gene of
              Candida boidinii
JOURNAL   Patent: US 6001590-A 41 14-DEC-1999;
FEATURES
SOURCE    1..38
          Location/Qualifiers
            1..38
            /organism="unknown"
            /db_xref="taxon:32644" 13 t
BASE COUNT 12 a 8 c 5 g
ORIGIN
Query Match 100.0%; Score 9; DB 9; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 10 TAATACTAG 18

RESULT 8
LOCUS      AX081631/c
DEFINITION Sequence 136 from Patent WO0109350.

```

Thu Oct 25 13:08:49 2001

us-09-462-955-4.ige

```

DB 10 TAATACTAG 18

RESULT 10
AX081623/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT
ORIGIN

AX081623
AX081631.1 GI:13170456
synthetic construct.
artificial sequence.
1 (bases 1 to 38)
Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
Genetically engineered bleb vaccine
Patent: WO 0109350-A 136 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
Location/Qualifiers
1..38
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="BAD20 primer"
7 a 14 c 6 g 11 t
11 t

Query Match 100.0%; Score 9; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
DB 26 TAATACTAG 18

RESULT 9
E22871
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Toshihiro YONEDA
DNA sequence increasing promoter activity
Patent: JP 199075843-A 8 23-MAR-1999;
KIRIN BREWERY CO LTD
OS Unidentified
PN JP 199075843-A/8
PD 23-MAR-1999
PF 29-AUG-1997 JP 1997234995
PR
PI
PC C12N15/09,C12N1/19,C12P21/02,((C12N15/09,C12R1:72),(C12N1/19,
C12R1:72)),
PC C12R1:72),
PC (C12P21/02,C12R1:72),C12N15/00,(C12N15/00,C12R1:72) CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..38
/organism="Unidentified".
Location/Qualifiers
1..38
/organism="unidentified"
/db_xref="taxon:32644"
12 a 8 c 5 g 13 t

Query Match 100.0%; Score 9; DB 10; Length 38;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||

```

```

DB 10 TAATACTAG 18

RESULT 10
AX081623/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT
ORIGIN

AX081623
AX081623.1 GI:13170448
synthetic construct.
artificial sequence.
1 (bases 1 to 39)
Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Feron,C.S.,
Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.
Genetically engineered bleb vaccine
Patent: WO 0109350-A 128 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
Location/Qualifiers
1..39
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="BAD 03-2 primer"
6 a 14 c 10 g 9 t

Query Match 100.0%; Score 9; DB 10; Length 39;
Best Local Similarity 100.0%; Pred. No. 8.7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
DB 27 TAATACTAG 19

RESULT 11
AX081623/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
BASE COUNT
ORIGIN

AX081623
AX081623.1 GI:12848534
Sequence 106 from patent US 6096321.
AR104937
AR104937
AR104937.1 GI:12848534
Unknown.
Unclassified.
1 (bases 1 to 41)
Girardeau,J., Martin,C., Mechin,M., Der Vartanian,M. and
Bousquet,F.
C1pG subunit of CS31A protein capsule containing heterologous
peptides
Patent: US 6096321-A 106 01-AUG-2000;
Location/Qualifiers
1..41
/organism="unknown"
11 a 3 c 10 g 17 t

Query Match 100.0%; Score 9; DB 9; Length 41;
Best Local Similarity 100.0%; Pred. No. 8.6e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
DB 9 TAATACTAG 1

RESULT 12
AX9423/c
LOCUS
DEFINITION
Sequence 42 from Patent WO9414967.
A39423
A39423
Sequence 42 from Patent WO9414967.
PAT 05-MAR-1997

```

```

ACCESSION   A39423
VERSION     A39423.1  GI:2295756
KEYWORDS
SOURCE      unidentified.
ORGANISM     unidentified.
REFERENCE    1 (bases 1 to 45)
AUTHORS      Girardeau, J., Martin, C., Mechin, M., Der, V.M. and Bousquet, F.
TITLE        CS31A PROTEIN CAPSULE SUBUNIT MODIFIED BY AT LEAST ONE HETEROLOGOUS
              PEPTIDE, CS31A PROTEIN CAPSULE COMPRISING SAID SUBUNIT,
              MICRO-ORGANISMS HAVING SUCH SUBUNITS ON THEIR OUTER MEMBRANE, AND
              METHODS FOR PREPARING AND USING SAME
JOURNAL      Patent: WO 9414967-A 42 07-JUL-1994;
              AGRONOMIQUE INST NAT RECH (FR)
COMMENT      Other publication AU 6808794 940719
              Other publication FR 2695538 940624.
FEATURES
  source     1..45
              /organism="unidentified"
              /db_xref="taxon:32644"
BASE COUNT   12 a 4 c 11 g 18 t
ORIGIN

Query Match      100.0%; Score 9; DB 9; Length 45;
Best Local Similarity 100.0%; Pred. No. 8.5e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taactag 9
    |||||
Db 9 TAATACTAG 1

RESULT 13
LOCUS      A39327/c
DEFINITION Sequence 9 from Patent WO9414969.
ACCESSION  A39327
VERSION    A39327.1  GI:2295680
KEYWORDS
SOURCE      unidentified.
ORGANISM     unidentified.
REFERENCE    1 (bases 1 to 62)
AUTHORS      Adams, S.E., Burns, R.N. and Richardson, S.M.
TITLE        NOVEL PROTEINACEOUS PARTICLES
JOURNAL      Patent: WO 9414969-A 9 07-JUL-1994;
              BRITISH BIO TECHNOLOGY (GB)
              Other publication AU 5711294 940719
              Other publication JP 8505768T 960625.
FEATURES
  source     1..62
              /organism="unidentified"
              /db_xref="taxon:32644"
BASE COUNT   23 a 12 c 10 g 17 t
ORIGIN

Query Match      100.0%; Score 9; DB 9; Length 62;
Best Local Similarity 100.0%; Pred. No. 7.9e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taactag 9
    |||||
Db 9 TAATACTAG 1

RESULT 14
LOCUS      AX069021/c
DEFINITION Sequence 1 from Patent WO0102555.
ACCESSION  AX069021
VERSION    AX069021.1  GI:12578836
KEYWORDS
SOURCE      unidentified.
ORGANISM     unidentified.
REFERENCE    1 (bases 1 to 74)
AUTHORS      Girardeau, J., Martin, C., Mechin, M., Der, V.M. and Bousquet, F.
TITLE        CS31A PROTEIN CAPSULE SUBUNIT MODIFIED BY AT LEAST ONE HETEROLOGOUS
              PEPTIDE, CS31A PROTEIN CAPSULE COMPRISING SAID SUBUNIT,
              MICRO-ORGANISMS HAVING SUCH SUBUNITS ON THEIR OUTER MEMBRANE, AND
              METHODS FOR PREPARING AND USING SAME
JOURNAL      Patent: WO 9414967-A 42 07-JUL-1994;
              AGRONOMIQUE INST NAT RECH (FR)
COMMENT      Other publication AU 6808794 940719
              Other publication FR 2695538 940624.
FEATURES
  source     1..74
              /organism="unidentified"
              /db_xref="taxon:32644"
BASE COUNT   12 a 4 c 11 g 18 t
ORIGIN

```

```

KEYWORDS
SOURCE      Mycobacterium tuberculosis.
ORGANISM     Mycobacterium tuberculosis.
              Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Actinomycetales; Corynebacterineae; Mycobacteriaceae;
              Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE    1 (bases 1 to 74)
AUTHORS      Gicquel, B., Guilhot, C. and Camacho, L.
TITLE        Method of making and identifying attenuated microorganisms,
              and preparations containing the sequences responsible for attenuation,
              and preparations containing attenuated microorganisms
JOURNAL      Patent: WO 0102555-A 1 11-JAN-2001;
              INSTITUT PASTEUR (FR)
FEATURES
  source     1..74
              Location/Qualifiers
              /organism="Mycobacterium tuberculosis"
              /db_xref="taxon:1773"
BASE COUNT   21 a 9 c 8 g 16 t 20 others
ORIGIN

Query Match      100.0%; Score 9; DB 10; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.6e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taactag 9
    |||||
Db 72 TAATACTAG 64

RESULT 15
LOCUS      AF227763/c
DEFINITION Hepatitis C virus isolate C10 envelope protein 2 (E2) gene, partial
              cds.
ACCESSION  AF227763
VERSION    AF227763.1  GI:8099446
KEYWORDS
SOURCE      Hepatitis C virus.
              Hepatitis C virus.
              Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
              Hepacivirus.
              1 (bases 1 to 81)
              Ross, R.S., Viazov, S., Gross, T., Hofmann, F., Seipp, H.M. and
              Roggendorf, M.
              Transmission of the hepatitis C virus from a patient to an
              anesthesiology assistant to five patients
              N. Engl. J. Med. 343 (25), 1851-1854 (2000)
              21004055
              PUBMED 11117977
              11117977
              2 (bases 1 to 81)
              Ross, S., Viazov, S. and Roggendorf, M.
              Direct Submission
              Submitted (21-JAN-2000) Institute of Virology, Essen University
              Hospital, Hufelandstr. 55, Essen D-45122, Germany
              Location/Qualifiers
              1..81
              /organism="Hepatitis C virus"
              /isolate="C10"
              /db_xref="taxon:11103"
              /note="genotype: 1a"
              <1..>81
              /gene="E2"
              <1..>81
              /gene="E2"
              /note="hypervariable region 1; HVRI"
              /codon_start=1
              /product="envelope protein 2"
              /protein_id="AAF72123.1"
              /db_xref="GI:8099447"
              /translation="HTHTGGSAAYTSSITRLTTPGAKQN"
BASE COUNT   21 a 29 c 18 g 13 t
ORIGIN

```

Thu Oct 25 13:08:49 2001

us-09-462-955-4.rge

Query Match 100.0%; Score 9; DB 58; Length 81;
Best Local Similarity 100.0%; Pred. No. 7.5e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
|||||||
Db 49 TAATACTAG 41

Search completed: October 24, 2001, 11:42:58
Job time: 6363 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:53 : Search time 5701.85 Seconds
(without alignments)
24.415 Million cell updates/sec

Title: US-09-462-955-3
Perfect score: 9
Sequence: 1 taatattac 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_hum1:*
34: em_hum2:*
35: em_hum3:*
36: em_hum4:*
37: em_hum5:*
38: em_hum6:*
39: em_hum7:*
40: em_in:*
41: em_om:*
42: em_or:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rol2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	9	100.0	17	10	I21591	I21591 Sequence 3
2	9	100.0	20	10	I19565	I19565 Sequence 3
3	9	100.0	20	10	I76521	I76521 Sequence 1
4	9	100.0	20	10	I76526	I76526 Sequence 6
5	9	100.0	21	9	AR102784	AR102784 Sequence
6	9	100.0	21	10	E23427	E23427 Oligonucleo
7	9	100.0	21	10	E23436	E23436 Oligonucleo
8	9	100.0	21	10	E23445	E23445 Oligonucleo

9 100.0 21 10 I76530
 10 100.0 24 10 E26401
 11 100.0 24 10 E26401
 12 100.0 26 10 AX073421
 13 100.0 27 9 AX027097
 14 100.0 30 97 WTHSNDU7M
 15 100.0 34 9 A49364
 16 100.0 35 9 A62540
 17 100.0 35 9 AR104880
 18 100.0 37 9 AR070909
 19 100.0 55 10 E22318
 20 100.0 58 58 BCTSTLPA
 21 100.0 60 9 A84543
 22 100.0 60 58 BCTSTLPP
 23 100.0 72 9 AR006900
 24 100.0 72 9 AR110924
 25 100.0 72 10 I74840
 26 100.0 75 6 DSNTSR4
 27 100.0 75 6 DSNTSR4
 28 100.0 76 7 TIOSINE
 29 100.0 76 7 TVU14602
 30 100.0 80 97 HSU14601
 31 100.0 80 97 HUMSINE
 32 100.0 90 3 STABLAZA
 33 100.0 95 5 AF166131
 34 100.0 98 5 AF166132
 35 100.0 98 54 G37851
 36 100.0 99 93 ISA241951
 37 100.0 99 93 HSA241951
 38 100.0 101 5 AF166133
 39 100.0 103 54 G32992
 40 100.0 104 92 HS87H7F
 41 100.0 105 53 CNS0740W
 42 100.0 109 54 G19822
 43 100.0 110 54 G37869
 44 100.0 110 58 BCU57490
 45 100.0 111 54 G38825

ALIGNMENTS

RESULT 1
 I21591 121591 17 bp DNA PAT 07-OCT-1996
 LOCUS Sequence 3 from patent US 5521302.
 DEFINITION I21591
 ACCESSION I21591
 VERSION I21591.1 GI:1601945
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 17)
 AUTHORS Cook,P.D.
 TITLE Process for preparing oligonucleotides having chiral phosphorus linkages
 JOURNAL Patent: US 5521302-A 3 28-MAY-1996;
 FEATURES Location/Qualifiers
 source 1. 17
 /organism="unknown"
 BASE COUNT 6 a 3 c 2 g 6 t
 ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 taatattac 9
 Db 7 TAATATTAC 15

RESULT 2
 I19565
 LOCUS I19565 20 bp DNA PAT 07-OCT-1996
 DEFINITION Sequence 3 from patent US 5506212.
 ACCESSION I19565
 VERSION I19565.1 GI:1599920
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 20)
 AUTHORS Hoke,G. and Cook,P.D.
 TITLE Oligonucleotides with substantially chirally pure phosphorothioate linkages
 JOURNAL Patent: US 5506212-A 3 09-APR-1996;
 FEATURES Location/Qualifiers
 source 1. 20
 /organism="unknown"
 BASE COUNT 6 a 3 c 3 g 8 t
 ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.7e-05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 taatattac 9
 Db 10 TAATATTAC 18

RESULT 3
 I76521 176521 20 bp DNA PAT 03-APR-1998
 LOCUS I76521
 DEFINITION Sequence 1 from patent US 5691461.
 ACCESSION I76521
 VERSION I76521.1 GI:3012675
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 20)
 AUTHORS Ecker,D.J. and Hoke,G.D.
 TITLE Oligonucleotides inhibiting candida germ tube formation
 JOURNAL Patent: US 5691461-A 1 25-NOV-1997;
 FEATURES Location/Qualifiers
 source 1. 20
 /organism="unknown"
 BASE COUNT 6 a 3 c 3 g 8 t
 ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.7e-05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 taatattac 9
 Db 10 TAATATTAC 18

RESULT 4
 I76526 176526 20 bp DNA PAT 03-APR-1998
 LOCUS I76526
 DEFINITION Sequence 6' from patent US 5691461.
 ACCESSION I76526
 VERSION I76526.1 GI:3012680
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 20)
 AUTHORS Ecker,D.J. and Hoke,G.D.

TITLE Oligonucleotides inhibiting candida germ tube formation
JOURNAL Patent: US 5691461-A 6 25-NOV-1997;

FEATURES
Source Location/Qualifiers
1..20
/organism="unknown"
BASE COUNT 6 a 3 c 3 g 8 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 10 TAATATTAC 18

RESULT 5
AR102784/c
LOCUS 21 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 11 from patent US 6087104.
ACCESSION AR102784
VERSION AR102784.1 GI:12814372
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 21)
AUTHORS Yamada,S., Venkateswaran,K. and Ohashi,E.
TITLE Oligonucleotides for detection of Bacillus cereus group bacteria harmful to mammals, and method of detection with the oligonucleotides
JOURNAL Patent: US 6087104-A 11 11-JUL-2000;
FEATURES Location/Qualifiers
Source 1..21
/organism="unknown"

BASE COUNT 9 a 3 c 3 g 6 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 13 TAATATTAC 5

RESULT 6
E23427/c
LOCUS 21 bp DNA PAT 07-FEB-2001
DEFINITION Oligonucleotide for detecting Bacillus cereus and method of detection therewith.
ACCESSION E23427
VERSION E23427.1 GI:13024403
KEYWORDS JP 1999004691-A/8.
SOURCE unidentified.
ORGANISM unclassified.

REFERENCE 1 (bases 1 to 21)
AUTHORS Shiochi,Y.K.V.V. and Ohashi.
TITLE Oligonucleotide for detecting Bacillus cereus and method of detection therewith
JOURNAL Patent: JP 1999004691-A 8 12-JAN-1999;
COMMENT NIPPON SUISAN KAISHA LTD
OS Unidentified
PN JP 1999004691-A/8
PD 12-JAN-1999
PF 29-SEP-1997 JP 1997/264057
PR
PI SHOICHI YAMADA, KASUTORI VENKATESWARAN,EIJI OHASHI PC

C12N15/09,C12Q1/04,C12Q1/68//((C12N15/09,C12R1:085),(C12Q1/04,PC C12R1:085)),
PC C12N15/00,(C12N15/00,C12R1:085)
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..21
FT /organism='Unidentified'.

FEATURES
source Location/Qualifiers
1..21
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 9 a 3 c 3 g 6 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 13 TAATATTAC 5

RESULT 7
E23436/c
LOCUS 21 bp DNA PAT 07-FEB-2001
DEFINITION Oligonucleotide for detecting Bacillus thuringiensis and method of detection therewith.
ACCESSION E23436
VERSION E23436.1 GI:13024412
KEYWORDS JP 1999004692-A/8.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 21)
AUTHORS Shiochi,Y.K.V.V. and Ohashi.
TITLE Oligonucleotide for detecting Bacillus thuringiensis and method of detection therewith
JOURNAL Patent: JP 1999004692-A 8 12-JAN-1999;
COMMENT NIPPON SUISAN KAISHA LTD
OS Unidentified
PN JP 1999004692-A/8
PD 12-JAN-1999
PF 29-SEP-1997 JP 1997/264058
PR

PI SHOICHI YAMADA, KASUTORI VENKATESWARAN,EIJI OHASHI PC
C12N15/09,C12Q1/68//((C12N15/09,C12R1:07),(C12N15/00,C12N15/00,PC C12R1:07))
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..21
FT /organism='Unidentified'.

FEATURES
source Location/Qualifiers
1..21
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 9 a 3 c 3 g 6 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 13 TAATATTAC 5

RESULT 8

```

E23445/c
LOCUS E23445 21 bp DNA PAT 07-FEB-2001
DEFINITION Oligonucleotide for detecting Bacillus anthracis and method of
detection thereof.
ACCESSION E23445
VERSION 1 GI:13024421
KEYWORDS JP 1999004693-A/8.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 21)
AUTHORS Shoichi,Y.K.V.V. and Ohashi.
TITLE Oligonucleotide for detecting Bacillus anthracis and method of
detection thereof.
JOURNAL NIPPON SUISEN KAISHA LTD
COMMENT Patent: JP 1999004693-A 8 12-JAN-1999;
JP 1999004693-A/8
PD 12-JAN-1999
PF 29-SEP-1997 JP 1997264059
PR SHOICHI YAMADA,KASUTORI VENKATESWARAN,EIJI OHASHI PC
C12N15/09,C12Q1/04,C12Q1/68//((C12N15/09,C12R1:07),(C12Q1/04,PC
C12R1:07)),
PC C12N15/00,(C12N15/00,C12R1:07)
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..21 /organism='Unidentified'.
FEATURES
source
1..21
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 9 a 3 c 3 g 6 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taatattac 9
Db 13 TAATATTAC 5
RESULT 9
LOCUS I76530 21 bp DNA PAT 03-APR-1998
DEFINITION Sequence 10 from patent US 5691461.
ACCESSION I76530
VERSION I76530.1 GI:3012684
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 21)
AUTHORS Ecker,D.J. and Hoke,G.D.
TITLE Oligonucleotides inhibiting candida germ tube formation
JOURNAL Patent: US 5691461-A 10 25-NOV-1997;
FEATURES
source
1..21
/organism='unknown'
BASE COUNT 8 a 4 c 2 g 7 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taatattac 9

```

```

E26401
LOCUS E26401 24 bp DNA PAT 07-FEB-2001
DEFINITION h-Hyd Protein, polynucleotide encoding said protein, antisense
polynucleotide thereof and antibody recognizing said protein.
ACCESSION E26401
VERSION E26401.1 GI:13026186
KEYWORDS JP 1999146786-A/2.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Naoki,N.K.N.N., Nakao,Y.H.M.M. and Hideyuki,S.
TITLE h-Hyd Protein, polynucleotide encoding said protein, antisense
polynucleotide thereof and antibody recognizing said protein
JOURNAL Patent: JP 1999146786-A 2 02-JUN-1999;
SUMITOMO ELECTRIC IND LTD
COMMENT OS Unidentified
PN JP 1999146786-A/2
PD 02-JUN-1999
PF 17-NOV-1997 JP 1997314935
PR NAOKI NIBARA,KYOKO NAGAMINE,MITSUYOSHI NAKAO,YOSHIOI HONDA,
PI MASAYUKI ANDO,
PI HIDEYUKI SAYA,
PC C12N15/09,C07K14/47,C07K16/18,C12P21/02//((C12N15/09,C12R1:91),
PC (C12P21/02,C12R1:19),C12N15/00,(C12N15/00,C12R1:91) CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..24 /organism='Unidentified'.
FEATURES
source
1..24
/organism='unidentified'
/db_xref='taxon:32644'
BASE COUNT 11 a 5 c 2 g 6 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taatattac 9
Db 4 TAATATTAC 12
RESULT 11
LOCUS E26401 24 bp DNA PAT 07-FEB-2001
DEFINITION h-Hyd Protein, polynucleotide encoding said protein, antisense
polynucleotide thereof and antibody recognizing said protein.
ACCESSION E26401
VERSION E26401.1 GI:13026186
KEYWORDS JP 1999146786-A/2.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Naoki,N.K.N.N., Nakao,Y.H.M.M. and Hideyuki,S.
TITLE h-Hyd Protein, polynucleotide encoding said protein, antisense
polynucleotide thereof and antibody recognizing said protein
JOURNAL Patent: JP 1999146786-A 2 02-JUN-1999;
SUMITOMO ELECTRIC IND LTD
COMMENT OS Unidentified
PN JP 1999146786-A/2
PD 02-JUN-1999

```

```

PF 17-NOV-1997 JP 1997314935
PR NAKI NIBARA,KYOKO NAGAMINE,MITSUYOSHI NAKAO,YOSHIOMI HONDA,
PI MASAYUKI ANDO,
PI HIDEYUKI SAVA,
PC C12N15/09,C07K14/47,C07K16/18,C12P21/02/(C12N15/09,C12R1:91),
PC (C12P21/02,C12R1:19),C12N15/00,(C12N15/00,C12R1:91) CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..24
FT /organism='Unidentified'.
FEATURES
    source
        Location/Qualifiers
            1..24
            /organism='unidentified'
            /db_xref='taxon:32644'
BASE COUNT 11 a 5 c 2 g 6 t
ORIGIN
Query Match 100.0%; Score 9; DB 10; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 11 TAATATTAC 3

RESULT 12
AX073421
LOCUS 26 bp DNA PAT 25-JAN-2001
DEFINITION Sequence 9 from Patent WO0102540.
ACCESSION AX073421
VERSION AX073421.1 GI:12583527
KEYWORDS
SOURCE synthetic construct.
ORGANISM
REFERENCE 1 (bases 1 to 26)
AUTHORS Hermiston,T., Hawkins,L.K. and Johnson,L.
TITLE Adenoviral vectors for treating disease
JOURNAL Patent: WO 0102540-A 9 11-JAN-2001;
ONX PHARMACEUTICALS, INC. (US)
FEATURES
    source
        Location/Qualifiers
            1..26
            /organism='synthetic construct'
            /db_xref='taxon:32630'
            /note='Oligonucleotide used to mutagenize the E3 region of
            Adenovirus 5.'
BASE COUNT 9 a 4 c 6 g 7 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 10 TAATATTAC 18

RESULT 13
AX027097
LOCUS 27-bp DNA PAT 16-SEP-2000
DEFINITION Sequence 12 from Patent DE19855469.
ACCESSION AX027097
VERSION AX027097.1 GI:10188107
KEYWORDS
SOURCE synthetic construct.
ORGANISM
    source
        Location/Qualifiers
            1..27
            /organism='synthetic construct'
            /db_xref='taxon:32630'
            /note='Primer_bind'
BASE COUNT 7 a 7 c 6 g 7 t
ORIGIN

Query Match 100.0%; Score 9; DB 10; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 10 TAATATTAC 18

RESULT 13
AX027097
LOCUS 27-bp DNA PAT 16-SEP-2000
DEFINITION Sequence 12 from Patent DE19855469.
ACCESSION AX027097
VERSION AX027097.1 GI:10188107
KEYWORDS
SOURCE synthetic construct.
ORGANISM
    source
        Location/Qualifiers
            1..27
            /organism='synthetic construct'
            /db_xref='taxon:32630'
            /note='Primer_bind'
BASE COUNT 7 a 7 c 6 g 7 t
ORIGIN

```

```

REFERENCE 1 (bases 1 to 27)
AUTHORS Esrich,M.
JOURNAL Patent: DE 19855469-A 12 15-JUN-2000;
ESRICH MICHAEL (DE)
FEATURES
    source
        Location/Qualifiers
            1..27
            /organism='synthetic construct'
            /db_xref='taxon:32630'
            /note='Primer_bind'
BASE COUNT 7 a 7 c 6 g 7 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 5 TAATATTAC 13

RESULT 14
MTHSNDU7M
LOCUS 30 bp DNA PRI 22-SEP-1995
DEFINITION H.sapiens mitochondrial NDUFV2 gene intron 7-.
ACCESSION X84435
VERSION X84435.1 GI:995673
KEYWORDS NADH ubiquinone oxidoreductase; NDUFV2 gene.
SOURCE human.
ORGANISM
    Mitochondrion Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 30)
AUTHORS De Coe,R.I.F.M.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1995) R.I.F.M. De Coe, Human Genetics, University
Hospital Nijmegen, PO Box 9101, 6500 HB Nijmegen, NETHERLANDS
REFERENCE 2 (bases 1 to 30)
AUTHORS de Coe,R., Buddiger,P., Smeets,H., Geurts van Kessel,A.,
Morgan-Hughes,J., Weghuis,D.O., Overhauser,J. and van Oost,B.
TITLE Molecular cloning and characterization of the active human
mitochondrial NADH:ubiquinone oxidoreductase 24-kDa gene (NDUFV2)
and its pseudogene
JOURNAL Genomics 26 (3), 461-466 (1995)
MEDLINE 95331780
COMMENT Sequence overlapping with that under the acc#M22538.
FEATURES
    source
        Location/Qualifiers
            1..30
            /organism='Homo sapiens'
            /organelle='mitochondrion'
            /db_xref='taxon:9606'
            /chromosome='18'
            /clone_lib='The chromosome specific human genome cosmid'
            /clone='120H-11'
            /map='p11.2-p11.31'
            1..30
            /gene='NDUFV2'
            1..30
            /gene='NDUFV2'
            /note='(-); gene for NADH-ubiquinone oxidoreductase'
            /number=7
            /evidence=experimental
BASE COUNT 9 a 3 c 2 g 16 t
ORIGIN

Query Match 100.0%; Score 9; DB 97; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 11 TAATATTAC 3

```

Db 10 TAATATTAC 18

RESULT 15
A49364/c
LOCUS A49364 34 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 45 from Patent WO9607745.
ACCESSION A49364
VERSION A49364.1 GI:2302843
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 34)
AUTHORS Gicquel, B., Lim, E. M., Portnoi, D., Berthet, F. and Timm, J.
TITLE MYCOBACTERIA FUNCTIONAL SCREENING AND/OR EXPRESSION VECTORS
JOURNAL Patent: WO 9607745-A 45 14-MAR-1996;
PASTEUR INSTITUT (FR)
COMMENT Other publication FR 2724183 960308.
FEATURES
source
1..34
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 12 a 2 c 3 g 17 t
ORIGIN

Query Match 100.0%; Score 9; DB 9; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 16 TAATATTAC 8

Search completed: October 24, 2001, 11:42:55
Job time: 6360 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:49 ; Search time 5701.85 Seconds
(without alignments)
84.096 Million cell updates/sec

Title: US-09-462-955-2

Perfect score: 31

Sequence: 1 agccgcgggggtaataactagccccgcggct 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_p11:*
13: gb_p12:*
14: gb_p13:*
15: gb_p14:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rod:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	31	100.0	1291	58	CFDCG
C 2	21.2	68.4	9294	1	AE006200
C 3	20.6	66.5	981	58	AB000926
C 4	20.6	66.5	989	58	AB000924
C 5	20.6	66.5	990	58	AB000923
C 6	20.6	66.5	209346	70	AC026373
C 7	20.2	65.2	901	1	AF005765
C 8	20.2	65.2	1122	1	AF004344
					M29963 Coconut fol
					AE006200 Pasteurel
					AB000926 Milk vetc
					AB000924 Milk vetc
					AB000923 Milk vetc
					AC026373 Homo sapi
					AF005765 Unculture
					AF004344 Unidentif

TITLE Direct Submission
JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
55108, USA

FEATURES Location/Qualifiers

source 1..9294
/organism="Pasteurella multocida"
/strain="PM70"
/db_xref="taxon:747"
complement(161..853)
/gene="PM1626"
complement(161..853)
/gene="PM1626"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03710.1"
/db_xref="GI:12722027"
/translation="MKTGLIGGMSPESTVYRILINOLINQAKSGNSAPLVMVNV
FEQICLOKQGEHKGAGEILAAKKLTOIGAEGILLATNTMHWKVAPOI IDVIDVPL
HIIDAVAKAIKINLTOVALLGTQFTMQDPFYRDILLHHGITPLVPNGKQOQEIRRI
FTELCVQIKPESQFYLDVINDLAKQAGGVILSGTEIGLLINQODATLFFDFDTTL
HAQAAAEFLTE"
complement(872..1102)
/gene="ilvM"
/note="PM1627"
complement(872..1102)
/gene="ilvM"
/codon_start=1
/transl_table=11
/product="ilvM"
/protein_id="AAK03711.1"
/db_xref="GI:12722028"
/translation="MOTYQFTLDVNRQPTLEIRLVRHGRFYVVMNMTAQEQAN
IDFTVKSDFRAEALLTNOLSKVVDLAIHQVR"
complement(1140..2789)
/gene="ilvG"
/note="PM1628"
complement(1140..2789)
/gene="ilvG"
/codon_start=1
/transl_table=11
/product="ilvG"
/protein_id="AAK03712.1"
/db_xref="GI:12722029"
/translation="MNGARLVTESLKTHGVTVTFYGPGLMPVVDALYDAGTEHLIC
RNEQGAAMAIAGYARSGKGVCIATSGGAANLITGLDALMDSIPVVALTQVAP
LIGTDAFOEDVLGSLACTKHSFLVQSIIEELPOVIAKAEIASGRPGVPLIDIPRD
IOIATDVTPTYSVAKPTALSOAHLDTALTLLKQAKRPVLYVGGVMAQAVPALRE
FLAVTQMSVSTLKLGSVHPDNPYMGIMGHGTCAANYAVQPCDILLAFGARFDDR
VTGKLDTPHAKYIHADIDMAETGKLRPPDVALCGDIVQAFNALSIPLNIAEWOAH
KOLKQTHDFCYSTNGDDTFINPLMLNSLSKKPKQSAVITTDVGHQHMWSAQHMOHYA
PENYITSAGYTGMFGPLPAACAKKARPDDDEVILISGDGSIIMNIOELGTFFKRGKTPV
KTVLLDNORLGMVQWQELFENAFSNTILDDNPDFVYMLAAAGIOGERITKGEVEEG
ADRLLEAEAFVFLHVCISSNENWPLVPVPPNACNLDMVEQM"
3149..3967
/gene="PM1629"
3149..3967
/gene="PM1629"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03713.1"
/db_xref="GI:12722030"
/translation="MSFPFRAIVSDLDGTLNNAHMGIDGTITQTLQOLAAKCIDIML
ATGRNHTDILLPILKKNVIERKAVMTSNGAQDLQGNLLVROYLPEQIAFDIMNLD
RORVSAVQCDWFENIDVPOLRYKHQDSGFMYEVDFQAHHGRETERKVFFIGREPO
DLGLEQLQANADTTSITYIPVCLFINNNKNSKSALEKVLADRDYDLQCLAFG
DGMNDVQLSRVGGVGNADPRLKEACPHELVIGNNAQESVANYIRTVEFIE"
3983..4387
/gene="crcB"
/note="PM1630"
3983..4387

gene
/gene="crcB"
/codon_start=1
/transl_table=11
/product="crcB"
/protein_id="AAK03714.1"
/db_xref="GI:12722031"
/translation="MAGSRSVISVGQOIIIFSSGAALGALSRWGLGLLLNPLFSAFSL
GTILVNYGLGIIGVFLAFWQYPOCSAEWRLLFFVTGFLGSLTTFSTSAEVIENLQ
QKWLAGLMLASGHLGLCLFTALGVFIWYQ"
complement(4365..5318)
/gene="pldB"
/note="PM1631"
complement(4365..5318)
/gene="pldB"
/codon_start=1
/transl_table=11
/product="pldB"
/protein_id="AAK03715.1"
/db_xref="GI:12722032"
/translation="MIREPHFTFALAEHPFAQFPQLQDVQGGKGVRTYRHFVQDN
PVORNLVIVNGRAENLKWTELAVDFYQGVYDVLVDFHGRGQYSORLLPOHEKGYID
EPFYADDMAILLAKISALYPVENOHILLASLIGALISTYYLANHDHQKSAVSAPPY
GIPLKHSDRELILNMLLGGQSRVYFGKGRYPADLDNLDLSCCRMRMMNRINR
RYPHILHGGPTFRWHLCESAINKLPTILPRIEIPVLILOSEKKEKIVENKNLRLTAL
LPREUQIAQSKHEILFPERDLIRSEALKKISQFTANSK"
complement(5460..6569)
/gene="asd"
/note="PM1632"
complement(5460..6569)
/gene="asd"
/codon_start=1
/transl_table=11
/product="asd"
/protein_id="AAK03716.1"
/db_xref="GI:12722033"
/translation="MONVGFIGWRGVSVLMDRMQENDEANINPVFTTSQAGQNA
PVGCKQAGELKDAFDIAELQKLDIIVTCQGGDYTNVYKPLRAAGKMGYWDIASAL
RMEKDAIIVLDPVNONVISEGLKNGIKTYGGNCTVSLMLMALGGLFERDLVEMVSA
TYAASAGAGAKNMRELLVOMGOLEDVSKVDLNDNPASSILDIERRKVTAKMRGDLPIENF
GAPLAGSLIPWIDKLWEDGOTKEWKGVAETNKILGDDSPIPVDGLCVRIGALRCHS
QAFITMKKDLPLIEIEQILASHNEWKVIIPNEKDVILRELTPTKVTGTLSVPVGRLR
KLKMGPEYLAAPTGVGDLQLLWGAAEPLRLILKQLVA"
6770..7009
/gene="PM1633"
6770..7009
/gene="PM1633"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03717.1"
/db_xref="GI:12722034"
/translation="MDFLCFVAFVAPRDERYQKRSVYVIOPTERKLGKASISLRNFS
SDTESKAWMTIDINIKIPEENKLFENKSTELKFPQ"
7028..7381
/gene="PM1634"
7028..7381
/gene="PM1634"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03718.1"
/db_xref="GI:12722035"
/translation="MNPNNKKKILEKSKGLPLFALLAEFEDYFGSVKDSDETKELFLS
FVLEMDHGELKFAIRKGFLEGSTEEQIDLFQAWPDHYDDEKLEYDIDHLLWITVAP
AGAYVICEDGYEWT"
7583..7948
/gene="PM1635"
7583..7948
/gene="PM1635"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03719.1"

```

/db_xref="GI:12722036"
/translation="MLEKSEAQYKRVENRGLVLDGLFANFQCYGDIRDKPEIKE
MFLMLADLMKLTGKLGKFLGSEIEQIDLFQAPWDHYDDKELEYDIDLHWI
TYAPAGAVWIDGIEDWI"
8272.8637
/gene="PM1636"
CDS
8272.8637
/gene="PM1636"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03720.1"
/db_xref="GI:12722037"
/translation="MLEKSEAQYKRVENRGLTLIGMFSSQVCYCYGDIRDKPEIKE
MFLMLADLMKLTGKLGKFLGSEIEQIDLFQAPWDHYDDKELEYDIDLHWI
VYAPAGAVWIDGIEWT"
8707.9144
/gene="PM1637"
8707.9144
/gene="PM1637"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03721.1"
/db_xref="GI:12722038"
/translation="MMDYRWSSRSPDERYQDRKEVNMLEKSEAQYKRVENRGL
ALDGLFASQVCGDIDQDKPEIKEMFELMADLMKLTGKLGKFLGSEIEQIDV
FQAPWDHYDDKELEYDIDLHWIHYAPAGAVWIDGIEWT"
BASE COUNT 2772 a 1733 c 1979 g 2810 t
ORIGIN

Query Match 68.4%; Score 21.2; DB 1; Length 9294;
Best Local Similarity 88.5%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gcggggtaataactagcccgagc 30
||||| ||||| ||||| ||||| |||||
Db 676 GCGGGGAATACTACCCACTCGGC 651

RESULT 3
AB000926/c 981 bp DNA circular VRL 30-OCT-1998
LOCUS Milk vetch dwarf virus genome segment 7 encoding viral
DEFINITION non-structural protein, complete sequence.
ACCESSION AB000926
VERSION 1 GI:3808185
KEYWORDS viral non-structural protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM Viruses; ssDNA viruses; Nanovirus.

REFERENCE
AUTHORS Sano,Y.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel:075-724-7764, Fax:075-724-7764)
2 (sites)
Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
99094638
MEDLINE
JOURNAL
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798663.
Sequence updated (29-Nov-1997).
FEATURES
source
1..981
/organism="milk vetch dwarf virus"
/virion
/db_xref="taxon:67585"
/chromosome="5"
/lab_host="Pisum sativum"
6..34
stem_loop 279..285
TATA_signal 376..828
CDS
/function="unknown"
/note="putative"
/codon_start=1
/product="viral non-structural protein"
/protein_id="BAA33984.1"
/db_xref="GI:3798660"
/translation="MSVNDYSLFDEATDELVSERKLIADVCHDDSDQVINVKVEDIF
CDMSDKVLCYKRYKRLDITLLGCRMKVITELKTPSERLSKSILOKKNMII
DGNLYLGLRMFFINQLINTCKWITRIEDVPICTLYHVNNTPIVDI"
polyA_signal 827..832
BASE COUNT 284 a 167 c 210 g 328 t

```

```

6..34
stem_loop 280..286
TATA_signal 346..723
CDS
/function="unknown"
/note="putative"
/codon_start=1
/product="viral non-structural protein"
/protein_id="BAA33986.1"
/db_xref="GI:3798664"
/translation="MEDFKQPKLSYGEIVOMKEODAFWSCYHFEFLRNEDVLCMC
RGRKLPAYKLPITAPIRWVLTIRAITDVRVDECKSCSHEISRDRYNPIKKEGLK
LYDSGNRYQVYSSSCNRDKSD"
polyA_signal 730..735
BASE COUNT 291 a 170 c 214 g 306 t
ORIGIN

Query Match 66.5%; Score 20.6; DB 58; Length 981;
Best Local Similarity 85.2%; Pred. No. 4.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 gcggggtaataactagcccgagc 31
||||| ||||| ||||| ||||| |||||
Db 31 GCGGGGTAATAAGCCCGGCCCT 5

RESULT 4
AB000924/c 989 bp DNA circular VRL 30-OCT-1998
LOCUS Milk vetch dwarf virus genome segment 5 encoding viral
DEFINITION non-structural protein, complete sequence.
ACCESSION AB000924
VERSION 1 GI:3808183
KEYWORDS viral non-structural protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM Viruses; ssDNA viruses; Nanovirus.

REFERENCE
AUTHORS Sano,Y.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel:075-724-7764, Fax:075-724-7764)
2 (sites)
Sano,Y., Wada,M., Hashimoto,Y., Matsumoto,T. and Kojima,M.
Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
99094638
MEDLINE
JOURNAL
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798659.
Sequence updated (29-Nov-1997).
FEATURES
source
1..989
/organism="milk vetch dwarf virus"
/virion
/db_xref="taxon:67585"
/chromosome="5"
/lab_host="Pisum sativum"
6..34
stem_loop 279..285
TATA_signal 376..828
CDS
/function="unknown"
/note="putative"
/codon_start=1
/product="viral non-structural protein"
/protein_id="BAA33984.1"
/db_xref="GI:3798660"
/translation="MSVNDYSLFDEATDELVSERKLIADVCHDDSDQVINVKVEDIF
CDMSDKVLCYKRYKRLDITLLGCRMKVITELKTPSERLSKSILOKKNMII
DGNLYLGLRMFFINQLINTCKWITRIEDVPICTLYHVNNTPIVDI"
polyA_signal 827..832
BASE COUNT 284 a 167 c 210 g 328 t

```


ORIGIN

Query Match 66.5%; Score 20.6; DB 58; Length 989;
 Best Local Similarity 85.2%; Pred. No. 4.6e-02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ggggggtaataactagcccccgcgcct 31
 |||||
 Db 31 GCGGGGTAATACTAAGCCCGCCCT 5

RESULT 5

AB000923/c

LOCUS AB000923
 DEFINITION Milk vetch dwarf virus genome segment 4 encoding viral non-structural protein, complete sequence.

ACCESSION AB000923.1 GI:3808182
 VERSION AB000923
 KEYWORDS viral non-structural protein.

SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
 ORGANISM Viruses; ssDNA viruses; Nanovirus.

REFERENCE 1 (bases 1 to 990)
 Sano, Y.

DIRECT SUBMISSION
 Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.

Yoshitaka Sano, Kyoto Institute of Technology, Department of Applied Biology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan (Tel: 075-724-7764, Fax: 075-724-7764)

REFERENCE 2 (sites)
 Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.

Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome

J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)

99094638
 On Oct 29, 1998 this sequence version replaced gi:3798657.

FEATURES
 Location/Qualifiers

1..990
 /organism="milk vetch dwarf virus"

/db_xref="taxon:67585"

/chromosome="segment 4"

/lab_host="Pisum sativum"

6..34
 /product="viral non-structural protein"

279..285
 /protein_id="BAA33983.1"

319..828
 /db_xref="GI:3798658"

/translation="MGLRYFAHLPLEKRIVHLOERKKEFLKAIEDSCRRHEA
 LIIDPSPALNSLSKFLTALSDVYGNQFNTRCLIRWKKDVPKVKGFDEQHKLK
 GSDMDLSCGLFIPDEEDDLTYEDGVIVRCSQLDLFKSLGIEIYIVVSRKCIWA
 PLSKEIVIK"

polyA_signal 820..825
 BASE COUNT 285 a 161 c 226 g 318 t

ORIGIN

Query Match 66.5%; Score 20.6; DB 58; Length 990;
 Best Local Similarity 85.2%; Pred. No. 4.6e-02;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ggggggtaataactagcccccgcgcct 31
 |||||
 Db 31 GCGGGGTAATACTAAGCCCGCCCT 5

RESULT 6

AC026373

LOCUS AC026373 209346 bp DNA HTG 15-SEP-2000

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Homo sapiens chromosome 12 clone RP11-353N14, WORKING DRAFT
 SEQUENCE, 41 unordered pieces.

AC026373

AC026373.17 GI:10121937

HTG: HTGS_PHASE1; HTGS_DRAFT.

human

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 209346)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Banks, T., Barbara, J.,

Benton, J., Blevins, M., Brown, E., Brown, M., Bryant, N.P., Bouch, J.,

Bowie, S., Burrell, C., Burrell, K.L., Byrd, N.C., Carton, T.F.,

Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,

Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,

Coyne, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,

Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,

Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,

Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,

Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,

Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,

Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,

Hernandez, J., Hernandez, O., Hodgson, A., Hoque, M., Holloway, C.,

Hollins, B., Homs, P., Howard, S., Huber, J., Hulyk, S., Hume, J.,

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S.,

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,

Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W.,

Louise, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,

Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,

Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M.,

Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,

Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,

Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,

Ogih, M., Okuwonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,

Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,

Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,

Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,

Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,

Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,

Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,

Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,

Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,

Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.

and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 209346)

Worley, K.C.

Direct Submission

Submitted (22-MAR-2000) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Sep 14, 2000 this sequence version replaced gi:10120361.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HATU

Center clone name: RP11-353N14

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-primer Bodypy: 23% of reads

Chemistry: Dye-terminator Big Dye: 75% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 177549 bases at least Q40

Consensus quality: 193666 bases at least Q30

2 (bases 1 to 804)
Oline,D.K., Schmidt,S.K. and Grant,M.C.
Direct Submission
Submitted (18-DEC-2000) Environmental, Population, and Organismic
Biology, University of Colorado at Boulder, Campus Box 334,
University of Colorado, Boulder, CO 80309, USA
Ponderosa pine forest soil isolate, USA: Front Range of Colorado,
Al site, elev. 1800m.m.
Location/Qualifiers
1. .804
/organism="uncultured Front Range soil crenarchaeote
FRA31B"
/isolate="FRA31B"
/db_xref="taxon:147503"
/note="Ponderosa pine forest soil isolate
?: USA: Front Range of Colorado, Al site, elev. 1800m."
<1. .>804
/product="16S ribosomal RNA"
205 a 188 c 254 g 156 t 1 others

BASE COUNT
ORIGIN

Query Match 64.5%; Score 20; DB 2; Length 804;
Best Local Similarity 82.1%; Pred. No. 8.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 agccgcgggggtaactagccccgcg 28
||||| ||||||| ||| |||||
Db 431 AGCCGCGCGGTAATACCGACCGGAG 458

RESULT 10
AF112330

LOCUS AF112330 519 bp DNA VRT 26-FEB-2001
DEFINITION Galaxias paucispondylus 16S ribosomal RNA gene, mitochondrial gene
for mitochondrial RNA, partial sequence.
ACCESSION AF112330
VERSION AF112330.1 GI:7688534
KEYWORDS
SOURCE Galaxias paucispondylus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Protacanthopterygii; Osmeriformes; Galaxiidae; Galaxias.
1 (bases 1 to 519)
Waters,J.M., Lopez,J.A. and Wallis,G.P.
Molecular phylogenetics and biogeography of galaxiid fishes
(Osteichthyes: Galaxiidae): dispersal, vicariance and the position
of Lepidogalaxias salamandroides
Syst. Biol. 49 (4), 777-795 (2000)
2 (bases 1 to 519)
Waters,J.M. and Wallis,G.P.
Direct Submission
Submitted (08-DEC-1998) Department of Zoology, University of Otago,
PO Box 56, Dunedin, Otago, New Zealand
Location/Qualifiers
1. .519
/organism="Galaxias paucispondylus"
/organelle="mitochondrion"
/db_xref="taxon:89560"
<1. .>519
/product="16S ribosomal RNA"
144 a 126 c 129 g 120 t

BASE COUNT
ORIGIN

Query Match 63.9%; Score 19.8; DB 8; Length 519;
Best Local Similarity 91.3%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ccgcgggggtaactagcccc 25
||||| ||||||| |||||||
b28 CCGCGGGGGAAGACTAGCCCC 310

AB000925/c					30-OCT-1998
LOCUS	977 bp	DNA	circular	VRL	
DEFINITION	Milk vetch dwarf virus genome segment 6 encoding viral				

non-structural protein, complete sequence.
 AB000925.1 GI:3808184
 viral non-structural protein.
 milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
 ORGANISM
 milk vetch dwarf virus
 Viruses: ssDNA viruses; Nanovirus.
 1 (bases 1 to 977)
 Sano, Y.
 Direct Submission
 Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
 Yoshitaka Sano, Kyoto Institute of Technology, Department of
 Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
 (Tel:075-724-7764, Fax:075-724-7764)
 2 (sites)
 Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.
 Sequences of ten circular ssDNA components associated with the milk
 vetch dwarf virus genome
 J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
 99094638
 On Oct 29, 1998 this sequence version replaced gi:3798661.
 Location/Qualifiers
 1..977
 /organism="milk vetch dwarf virus"
 /viral
 /db_xref="taxon:67585"
 /chromosome="segment 6"
 /lab_host="Pisum sativum"
 1..33
 304..310
 363..824
 /function="unknown"
 /note="putative"
 /codon_start=1
 /product="viral non-structural protein"
 /protein_id="BAA33985.1"
 /db_xref="GI:3798662"
 /translation="MADWFAFLKTKTHVDFPTLACDPSOEITCCDSMNKLNDSRK
 VLLVSCVSFNQSGVGNVNRVGOLOISMLEDGVCVRPIGVPIGGYLYHNDYGYEG
 EKTFLNLDIESQVLKDEYDKRKFVSVNLNGLSLCDLKVFIHVHAIKRV"
 polyA_signal 823..828
 BASE COUNT 275 a 162 c 213 g 327 t
 ORIGIN

Query Match 63.9%; Score 19.8; DB 58; Length 977;
 Best Local Similarity 91.3%; Pred. No. 9.6e-02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 5 gcgggggtaatactagccccgc 27
 |||||
 Db 28 GCGGGGTAATACTAAGCCCCGC 6

RESULT 15
 AB000927/c
 LOCUS
 DEFINITION
 Milk vetch dwarf virus genome segment 8 encoding viral
 non-structural protein, complete sequence.
 AB000927
 VERSION
 AB000927.1 GI:3808186
 viral non-structural protein.
 Keywords
 milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
 SOURCE
 milk vetch dwarf virus
 Viruses: ssDNA viruses; Nanovirus.
 1 (bases 1 to 985)
 Sano, Y.
 Direct Submission
 Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
 Yoshitaka Sano, Kyoto Institute of Technology, Department of
 Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
 (Tel:075-724-7764, Fax:075-724-7764)
 2 (sites)

AUTHORS
 TITLE
 Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.
 Sequences of ten circular ssDNA components associated with the milk
 vetch dwarf virus genome
 J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
 JOURNAL
 MEDLINE
 99094638
 COMMENT
 On Oct 29, 1998 this sequence version replaced gi:3798665.
 FEATURES
 Location/Qualifiers
 1..985
 /organism="milk vetch dwarf virus"
 /viral
 /db_xref="taxon:67585"
 /chromosome="segment 8"
 /lab_host="Pisum sativum"
 1..33
 276..282
 344..682
 /function="unknown"
 /note="putative"
 /codon_start=1
 /product="viral non-structural protein"
 /protein_id="BAA33987.1"
 /db_xref="GI:3798666"
 /translation="MADPVYGYODGDIDAKRHOALYIGIILIMVCIILWVC
 IMLACYIPGFLKKTMEAWLSSSSMKRVAATITRTPEATGPERNWDAROTNAA
 SSOPNSGGVF"
 polyA_signal 775..780
 BASE COUNT 283 a 180 c 213 g 309 t
 ORIGIN

Query Match 63.9%; Score 19.8; DB 58; Length 985;
 Best Local Similarity 91.3%; Pred. No. 9.5e-02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 5 gcgggggtaatactagccccgc 27
 |||||
 Db 28 GCGGGGTAATACTAAGCCCCGC 6

Search completed: October 24, 2001, 11:42:53
 Job time: 6358 sec

us-09-462-955-2.rge

Thu Oct 25 13:08:39 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:48 ; Search time 5701.85 Seconds
(without alignments)
762.285 Million cell updates/sec

Title: US-09-462-955-1-copy_711_991

Perfect score: 281

Sequence: 1 cacatgtgtgtgaaccagaa.....ttaagtgatgtgtcatctaa 281

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_in1.*
- 5: gb_in2.*
- 6: gb_in3.*
- 7: gb_om.*
- 8: gb_ov.*
- 9: gb_pat1.*
- 10: gb_pat2.*
- 11: gb_ph.*
- 12: gb_pl1.*
- 13: gb_pl2.*
- 14: gb_pl3.*
- 15: gb_pl4.*
- 16: em_ba1.*
- 17: em_ba2.*
- 18: em_fun.*
- 19: em_htgo_hum.*
- 20: em_htgo_inv.*
- 21: em_htgo_rod.*
- 22: em_htg_hum1.*
- 23: em_htg_hum2.*
- 24: em_htg_hum3.*
- 25: em_htg_hum4.*
- 26: em_htg_hum5.*
- 27: em_htg_hum6.*
- 28: em_htg_hum7.*
- 29: em_htg_hum8.*
- 30: em_htg_inv1.*
- 31: em_htg_inv2.*
- 32: em_htg_other.*
- 33: em_htg_rod.*
- 34: em_hum1.*
- 35: em_hum2.*
- 36: em_hum3.*
- 37: em_hum4.*
- 38: em_hum5.*
- 39: em_hum6.*
- 40: em_hum7.*
- 41: em_in.*
- 42: em_om.*
- 43: em_or.*

44: em_ov.*

45: em_pat.*

46: em_ph.*

47: em_pl.*

48: em_ro.*

49: em_sts.*

50: em_sy.*

51: em_un.*

52: em_vi.*

53: gb_sts1.*

54: gb_sts2.*

55: gb_sts3.*

56: gb_sy.*

57: gb_un.*

58: gb_vil.*

59: gb_vil2.*

60: gb_htg1.*

61: gb_htg2.*

62: gb_htg3.*

63: gb_htg4.*

64: gb_htg5.*

65: gb_htg6.*

66: gb_htg7.*

67: gb_htg8.*

68: gb_htg9.*

69: gb_htg10.*

70: gb_htg11.*

71: gb_htg12.*

72: gb_htg13.*

73: gb_htg14.*

74: gb_htg15.*

75: gb_htg16.*

76: gb_htg17.*

77: gb_htg18.*

78: gb_htg19.*

79: gb_htg20.*

80: gb_htg21.*

81: gb_htg22.*

82: gb_htg23.*

83: gb_htg24.*

84: gb_htg25.*

85: gb_pr1.*

86: gb_pr2.*

87: gb_pr3.*

88: gb_pr4.*

89: gb_pr5.*

90: gb_pr6.*

91: gb_pr7.*

92: gb_pr8.*

93: gb_pr9.*

94: gb_rol.*

95: gb_rol2.*

96: gb_in4.*

97: gb_pr10.*

98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	281	100.0	1291	58	CFDCG
2	89.4	31.8	1007	58	M29963 Coconut fol
3	86.2	30.7	1004	59	AJ132187 faba bean
4	74	26.3	1007	58	AJ005966 faba bean
5	70.8	25.2	1091	9	AB000920 Milk vetc
6	70.8	25.2	1096	9	AR063453 Sequence
7	70.8	25.2	1096	9	AR063452 Sequence
8	70.8	25.2	1106	58	BYTV2
					L32167 Banana bunc
					AR063451 Sequence

```

TRGVKRPRIAQRAEPEDEPLEDPGGYRRCVHGA SVEWTRWAAENPFFFPYHNQ
LEVLASGEADPDORTILWTCGGDDGKSVFAKYLGLKPDWPFYTCGGTRKDVLYQYIE
DPKRNLLDVPACNLEVLNVALLECYNKRAFSDDRYEPLSYLGFEDHVLVLFANVLPD
YIKISDRRIKLWNI"
314..775
/note="orf 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSREVVTNRIETVRRNGCFSTESRLVLESKGHDWPN
DLLLNMSAWKQCADTEDALYTELRWNGDGPLKIRSHFHTIGSLKCCURSESQR
IAQSSGVADETEETGSGPCLNIDSSTGTSTHVVEPTYCTSTRTONEI"
complement(422..568)
/note="orf 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFQPIILSIPPKLRVQRFIRGLPLGGVHQVPPQIIVGP
IVAF"
639..797
/note="orf 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MRTRRRRREVVRVQCIISRTQARLVLMWNWQKGRIVPVHRGPKTK
FNPRCTQV"
complement(823..987)
/note="orf 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MHTLNLIPQYSVPADFOIIRQDIDGKEYYMHMVEPKITKGFEVFE
RTECPVLNTF"
1098..1286
/note="orf 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRWVGFTIKDSIWRITNLLGLQCTQPLSTSPIQVSSLLLEKK
AASLYLPSICFCAIGRLS"
336 a 323 c 332 g 300 t

```

```

BASE COUNT      336 a      332 c      332 g      300 t
ASLIDPSLCFARIGLS
ORIGIN
Query Match      100.0%; Score 281; DB 58; Length 1291;
Best local Similarity 100.0%; Pred. No. 9.6e-67;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cacatgtggtggaaccgaaagacgtattgtaccagttacatcagagaccacaaacgaaa 60
    |||
Db 711 CACATGTGTGGACACCAAGGACGTATTGTACCAAGTACATCGAGAGCCCAACACGAAA 770
    |||
QY 61 tttaatcctcgtatgccagggtgtaatttagagtgatttaaattatgcctctgtagaatg 120
    |||
Db 771 TTTAATCCTCCATGTACCCAGGTCTAATTTAGAGTATTTAAATATATGCCCTGTGAATG 830
    |||
QY 121 tgtaagaacaggggcatttcagtcggacaaatcacgaaccccttagttatcttgggttcga 180
    |||
Db 831 TGTTAAGAACACAGGGCATTCAGTTCGGACAAATAGCAACCCCTTAGTTCTTGGGTTGCA 890
    |||
QY 181 ccattgtcatgtactcgtatttgccaatgctcctgcctgattatttgaataatcagcaggaa 240
    |||
Db 891 CCATGTGCAATGTACTCGTATTGTCATGCTCGCTGATTATTTGAAAATCAGCAGGGA 950
    |||
QY 241 cagaataaacctgtggaatatattaaagtatgtgctaatctaa 281
    |||
Db 951 CAGAATAAACCTGTGGAATATTTAAAGATATGTGTCATCTAA 991
    |||
RESULT 2
FBE132187      1007 bp      DNA      circular      VRL      16-SEP-2000
LOCUS
DEFINITION     faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian

```

ALIGNMENTS

```

RESULT      1
CFDCG       1291 bp ss-DNA   circular    VRL          02-AUG-1993
LOCUS       foliar decay virus, complete genome.
DEFINITION  M29963
ACCESSION   M29963.1 GI:323306
VERSION     circular; complete genome.
KEYWORDS    Coconut foliar decay virus DNA, clones NMORG, PORG, and PORGSAU3A.
SOURCE      Coconut foliar decay virus
            Viruses; ssDNA viruses; Nanovirus.
ORIGINATOR  1 (bases 1 to 1291)
REFERENCE   Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
AUTHORS     Draft entry and printed sequence for [1] kindly submitted by
TITLE       W.Rhode, 15-MAR-1989, for release after publication.
JOURNAL     Location/Qualifiers
MEDLINE     1. .1291
COMMENT     /organism="Coconut:foliar decay virus"
            /db_xref="taxon:12474"
FEATURES    40_70
            /note="stem-loop structure"
            stem_loop 103..975
            CDS        /note="ORF 1"
                   /codon_start=1
                   /protein_id="AAA42894.1"
                   /db_xref="GI:323307"
                   /translation="MGSSLRRCWCTLNYETEEAAVVRRIESNLVYAIIVGDEVAPS

```


[illegible]

RESULT	3
LOCUS	NYV5966
DEFINITION	faba bean necrotic yellows virus C9 gene.
ACCESSION	AJ005966
VERSION	AJ005966.1 GI:3550532
KEYWORDS	C9 gene; component 9; putative; rep protein; replication associated protein.
SOURCE	faba bean necrotic yellows virus.
ORGANISM	faba bean necrotic yellows virus
REFERENCE	Viruses; ssDNA viruses; Nanovirus.
AUTHORS	1 (bases 1 to 1004)
TITLE	Katul, L.
	Direct Submission

pflanzenvirologie, Biologische Bundesanstalt fuer Land- und
 Forstwirtschaft, Messelweg 11 - 12, 38104 Braunschweig, GERMANY
 2 (bases 1 to 1004)
 Katul, L., Timchenko, T., Gronenborn, B. and Vetter, H. J.
 Ten distinct circular ssDNA components, four of which encode
 putative replication-associated proteins, are associated with the
 faba bean necrotic yellows virus genome
 J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
 99094637 Location/Qualifiers
 1..1004
 source
 FEATURES

```

/Virion
/isolate="SV292-88"
/db_xref="taxon:59817"
73..918
/gene="C9"
73..918
/gene="C9"
/function="putative replication associated (rep) protein"
/codon_start=1
/product="component 9"
/protein_id="CAA06789.1"

```

```

/obj_xref="G1:3550533"
/db_xref="SPTRMBL:O91252"
/translation="MSAVNWVFTLNFAGEVPVLSFDRVOYVWQHVRVNDHIQVY
QLKKAKNTYKNIIGGNPHLEKWKSGIEASAKVEGPMWSYGLLKGKSH
KXIMELIKDPNELIEPOKYRRAMAMDESRLAKEEGPYMFY"SWQETVGLLGE
EPNDRTIWTGPNCEKSGFGKFLGILKKDYLVLCGCKTQDMTYMLMKPKRANVYD
IYPCNSFEYVNPQELITGVATKSVKVRGCTVPLGK"

```

Query Match	30.7%;	Score 86.2;	DB:59;	Length 1004;
Best Local Similarity	58.3%;	Pred. No. 2e-13;		
Matches 151;	Conservative 0;	Mismatches 108;	Indels 0;	Gaps 0;

Y	7	tggtggaaaccagaaaggacgtattgtaccagtcacatogagaccacaaacgaatttaatt	66
b	660	TGGAGGTAAACCCCAAGATATGACATATATGTTATGAAAAATCCCAAGGCAATGT	719


```

ACCESSION   AR063452
VERSION     AR063452.1  GI:5992760
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1096)
AUTHORS     Wu, R., You, L., and Soong, T.
TITLE       Nucleotide sequence of two circular ssDNA associated with banana
            bunchy top virus and method for detection of banana bunchy top
            virus
JOURNAL     Patent: US 5846705-A 4 08-DEC-1998;
FEATURES    Location/Qualifiers
             source          1..1096
BASE COUNT   347 a 231 c 244 g 274 t
ORIGIN
1
Query Match      25.2%; Score 70.8; DB 9; Length 1096;
Best Local Similarity 56.2%; Pred. No. 3.4e-09;
Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

Qy 8 ggtggaaccagaagcgtatgtaccagtagacatcagagaccacaaacgaatttaac 67
Dy 668 GGTGGAAGACGTCGGATATGATGCACATCATACACGATGATCCTGATAATCATTTGGATT 727

Qy 68 ctcgatgtaccagtgtaatttagagtagatttaaaattatgccctgttagaattgttaag 127
Dy 728 ATTGATATCCCAAGTCATTCAGATTATTCGAATTATTCGGCTGTATAGAACAAATTAAG 787

Qy 128 aacaggcattcagttcgagacaaatcgaacc---ccttagttatcttgggttcgacct 184
Dy 788 AATAGATTTTAAATAATACAAATACGAACCATGTGTGATTAGAAAGATGGACAAAT 847

Qy 185 gtgcgtactgtatttgcgaatgtccctgcctgattatttgaataatcagcaggagacaga 244
Dy 848 GTCCATGTAAATTTATGCAAAATGTCCTGCTGATATTTGTAATAATTCAGAAGATAGA 907

Qy 245 ataaactgtggaattttaagtagtgcac 278
Dy 908 ATAAAAATAATTAATTTGTGAGAAAGGAAACTTC 941

RESULT 7
LOCUS      BYTV2 1096 bp. DNA VRL 30-OCT-1994
DEFINITION Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes,
            complete cds's.
ACCESSION   L32167
VERSION     L32167.1  GI:520791
KEYWORDS    stem loop.
SOURCE      Banana bunchy top virus DNA.
ORGANISM    Banana bunchy top virus
            Viruses; ssDNA viruses; Circoviridae.
REFERENCE   1 (bases 1 to 1096)
AUTHORS     Wu, R.-Y., You, L.-R., and Soong, T.-S.
TITLE       Nucleotide sequences of two circular single-stranded DNAs
            associated with banana bunchy top virus
JOURNAL     Unpublished (1994)
FEATURES    Location/Qualifiers
             source          1..1096
             /organism="Banana bunchy top virus"
             /db_xref="taxon:12585"
TATA_signal 1..7
stem_loop   71..928
gene        71..928
CDS         71..928
            /gene="V2"
            /gene="V2"
            /codon_start=1
            /protein_id="AAA51426.1"
            /db_xref="GI:520792"

```

```

/translation="MSSFKWCFTLYSSAAEREDFLALLKEEDVHYSVVGDEVAPATG
OKHOGYLSLKKIRLGGILKKYKSRHWEIAKGSDEQNRRCYSKETLVLEIGTPVP
GSKRKLDRFRESPELKMDDPSKYRCLAVESKIDARINSEWHELKQNKLIQH
IEGPDORSIIWYGPNGEGKSTFARYLSLPGWGYINGKGTSDMMHIITMDPNHW
IIDPRSHSDYLVNGVIEQIKNRVLINTRYEPCVIRKQGNVHVIVMANVLPDYCKIS
EDRIKIINC"
85..309
/gene="V1"
85..309
/gene="V1"
/codon_start=1
/protein_id="AAA51427.1"
/db_xref="GI:520793"
/translation="MVLHSELFLRCARLRSRSGGCSLLCRRRRSRSGHRPEAPP
GISIPEKINSRRIEKEVWLPQSLGDCEK"
complement(544..699)
/gene="C2"
complement(544..699)
/gene="C2"
complement(544..699)
/gene="C2"
/codon_start=1
/protein_id="AAA51428.1"
/db_xref="GI:520794"
/translation="MMCIIISDVFPPLIYPHGFNDRLAKVDFPSPPLGPVTQMLRS
SGTPTSMC"
complement(883..1059)
/gene="C1"
complement(883..1059)
/gene="C1"
complement(883..1059)
/codon_start=1
/protein_id="AAA51429.1"
/db_xref="GI:520795"
/translation="MGLRPNKRFINYNDFLMHLMIACSASRYVSGTCTFLIACGSGFLS
QQLIIFILSSEILQ"
BASE COUNT 347 a 231 c 244 g 274 t
ORIGIN
1
Query Match      25.2%; Score 70.8; DB 58; Length 1096;
Best Local Similarity 56.2%; Pred. No. 3.4e-09;
Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

Qy 8 ggtggaaccagaagcgtatgtaccagtagacatcagagaccacaaacgaatttaac 67
Dy 668 GGTGGAAGACGTCGGATATGATGCACATCATACGATGATCCTGATAATCATTTGGATT 727

Qy 68 ctcgatgtaccagtgtaatttagagtagatttaaaattatgccctgttagaattgttaag 127
Dy 728 ATTGATATCCCAAGTCATTCAGATTATTCGAATTATTCGGCTGTATAGAACAAATTAAG 787

Qy 128 aacaggcattcagttcgagacaaatcgaacc---ccttagttatcttgggttcgacct 184
Dy 788 AATAGATTTTAAATAATACAAATACGAACCATGTGTGATTAGAAAGATGGACAAAT 847

Qy 185 gtgcgtactgtatttgcgaatgtccctgcctgattatttgaataatcagcaggagacaga 244
Dy 848 GTCCATGTAAATTTATGCAAAATGTCCTGCTGATATTTGTAATAATTCAGAAGATAGA 907

Qy 245 ataaactgtggaattttaagtagtgcac 278
Dy 908 ATAAAAATAATTAATTTGTGAGAAAGGAAACTTC 941

RESULT 8
LOCUS      AR063451 1106 bp. DNA PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5846705.
ACCESSION   AR063451
VERSION     AR063451.1  GI:5992759
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1106)

```

[illegible][illegible]

RESULT	10			
AF216222	DNA	1095 bp	VRL	08-MAR-2001
LOCUS	Banana bunchy top virus satellite S2 replication initiation protein (ORF V1) gene, complete cds.			
DEFINITION	AF216222			
ACCESSION	AF216222.1	GI:12004327		
VERSION				
KEYWORDS	Banana bunchy top virus. Banana bunchy top virus Viruses; ssDNA viruses; Circoviridae. 1 (bases 1 to 1095)			
ORGANISM	Horser,C.L., Karam,M., Harding,R.M. and Dale,J.L. Additional rep-encoding DNAs associated with banana bunchy top virus			
REFERENCE	Arch. Virol. 146 (1), 71-86 (2001)			
AUTHORS	21163975			
TITLE	2 (bases 1 to 1095)			
JOURNAL				
MEDLINE				
REFERENCE				

(Tel:075-724-7764, Fax:075-724-7764)

REFERENCE 2 (sites)
AUTHORS Sano.Y., Wada.M., Hashimoto.Y., Matsumoto.T. and Kojima.M.
TITLE Sequences of ten circular ssDNA components associated with the milk
vetch dwarf virus genome
J.Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
MEDLINE 99094638
COMMENT On Oct 29, 1998 this sequence version replaced gi:3798653.
FEATURES Location/Qualifiers
source 1..1009
/organism="milk vetch dwarf virus"
/viral
/db_xref="taxon:67585"
/chromosome="segment 2"
/lab_host="Pisum sativum"
stem_loop 1..32
CDS 69..917
/function="helicase"
/codon_start=1
/product="viral replication-associated protein"
/protein_id="BAA33981.1"
/db_xref="GI:3798654"
/translation="MASKRPFCLNYKTALERETFLSRDELNVFCGEIAPITG
QKHLGVSMKLRGLGLKKKFGSIAHWIAKDDQNDYCTKELTAEICAPYVK
GNSRRKIMEIYEDPEEMKURDPDIALCKAKKLRBEYCSVESILRPMQIELHRL
MEVPDDRTIWAYDGGEGKSTFAPKELIKYGMFTAGGKTQDILLVYAGDPENRTAF
DVPRCSSEMMNYQAMENMKRNVFASIKYRPVDLCIRKKVHLIVFANVAPDPTKLSER
IVIINC"

polyA_signal 917..922
TATA_signal 930..936
BASE COUNT 318 a 174 c 235 g 282 t
ORIGIN

Query Match 24.6%; Score 69.2; DB 58; Length 1009;
Best Local Similarity 54.3%; Pred. No. 9.2e-09;
Matches 140; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 8 ggtggaaccagaaagacgtattgtaccagatcacatcgagaccacaaacgaatttaac 67
Db 660 GGAGGAAGAACCCAGACATCTGTATGTATGCGCAAGATCTTCGAAGAANTATTCG 719
QY 68 ctgcatgtacccaggtgttaatttagagattattaaattatgcctgttagaatgtgtta 127
Db 720 TTGCATGTACCCAGGTGTTCTTCGGAGATGATGAACATCAAGCAATGAAATGATGA 779
QY 128 aacaggccattcagttcggaacaaatagacacccttagttcttgggttcgaccatgtg 187
Db 780 AATAGAGATATTGCCAAGTACAAAATATAGCCCTGTAGATCTTTGTATAGAAAGAGTT 839
QY 188 catgtactcgtatttgccaatgtcctgcctgattattgaaatcagcaggagacgaata 247
Db 840 CATTTAATTGTATTGCCACGTGGCACCCTGACCCACAAAATTAAGTGAGGACAGATT 899
QY 248 aaactgtggaattatttaa 265
Db 900 GTAATTATCAATTGTTGA 917

RESULT 14
BBU12587 BBU12587 1127 bp DNA circular VRL 01-FEB-1995
LOCUS Banana bunchy top virus DNA IV ORF V1 and ORF C1 genes, complete
DEFINITION cds.
ACCESSION U12587
VERSION U12587
KEYWORDS U12587.1 GI:642393
SOURCE Banana bunchy top virus.
ORGANISM Banana bunchy top virus.
REFERENCES 1 (bases 1 to 1127)
AUTHORS Viruses; ssDNA viruses; Circoviridae.
TITLE Wu,R.-Y. and You,L.-R.
Nucleotide sequences of DNA III and DNA IV associated with banana

dsDNA component of banana bunchy top virus
Virology 198, 645-652 (1994)
REFERENCE 2 (bases 1 to 1095)
AUTHORS Sano.Y.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1993) Yu-Chan Chao, Institute of Molecular
Biology, 120, Sec.2., Yen-joe-yuan Rd., Nan-Kang, Taipei, Taiwan
115, Republic of China
FEATURES Location/Qualifiers
source 1..1095
/organism="Banana bunchy top virus"
/specific_host="Musa acuminata Colla"
/db_xref="taxon:12585"
/clone="pBTD18"
/notes="component 2 (BBTV-C2)"
82..390
/note="orf1"
/codon_start=1
/protein_id="AAAL17783.1"
/db_xref="GI:475434"
/translation="MSPSLKWCFTLNYSAARENFLSLKKEEDVHYAVVGDEVAPA
TGQKHLOGYLSLKKKWPRTIEEVLPCSLGDCQRRRREFEVLFORNPRIRSSCC
"
CDS 419..634
/note="orf2"
/codon_start=1
/protein_id="AAAL17784.1"
/db_xref="GI:475435"
/translation="MIARSPDRMKIEOEIFHRYQSVNKLKPKFEFVHPCLDRPMQI
QLTEAIDPEPDRSLINWYGPYGNKGK"
BASE COUNT 337 a 217 c 253 g 288 t
ORIGIN

Query Match 24.8%; Score 69.8; DB 59; Length 1095;
Best Local Similarity 56.2%; Pred. No. 6.3e-09;
Matches 131; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 20 aaggacgtattgtaccagtacatcgaggaccacaaacgaatttaactcctcgatgacc 79
Db 695 AATATCTTATTCTCCATGTGGACGAAGGATCTGACAGCATATAGTATTGATATTCCT 754
QY 80 aggtgtaattagagtatttaattatgcctgttgtagaattgttgaagaacaggccattc 139
Db 755 CGTGTGAATCAGGATATTAAATTATGATGTAATAGAGGCCATTAAGGATAGGCTTATA 814
QY 140 agtcggacaaatacagaacccttagttcttgggttcgaccatgtcgtactcgtga 199
Db 815 GAGAGTACTAAATACAAACCCATAAGATAGTGTGAATTAGGTAAATACATCACTAATGTC 874
QY 200 ttgcgaatgctcgtcgtatttattgaaatcagcaggagacagaataaaact 252
Db 875 ATGCGAATTTTCATGCTGCTTCTTAAATCTCGAAGATCGAATAAAAAAT 927

RESULT 13
AB000921 1009 bp DNA circular VRL 30-OCT-1998
LOCUS Milk vetch dwarf virus genome segment 2 encoding viral
DEFINITION replication-associated protein, complete sequence.
ACCESSION AB000921.1 GI:3808180
VERSION
KEYWORDS viral replication-associated protein.
SOURCE milk vetch dwarf virus (lab_host:Pisum sativum) DNA.
ORGANISM milk vetch dwarf virus
Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1009)
AUTHORS Sano.Y.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology; Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:47 ; Search time 5701.85 Seconds
(without alignments)
1033.561 Million cell updates/sec

Title: US-09-462-955-1_COPY_611_991

Perfect score: 381

Sequence: 1 cagcggacgatcgacaatc.....ttaaagtatgtgtcatctaa 381

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vi.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vil.*
59: gb_vil2.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_rol.*
95: gb_rol2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	381	100.0	1291	58	CFDCG
2	115.8	30.4	1007	58	M29963 Coconut fol
3	114.2	30.0	1004	59	AJ132187 faba bean
4	112.4	29.5	1096	9	AJ005966 faba bean
5	112.4	29.5	1096	58	AR063452 Sequence
6	110.8	29.1	1091	9	L32167 Banana bunc
7	110.6	29.0	1095	58	AR063453 Sequence
8	96	25.2	1111	58	AF216222 Banana bu
					U12586 Banana bunc

	TRPVKRPRLQAFAEPEDELRLLEDPGGYRRCVVHCASVEWTRWAENPFPFPPYHNC
	LEVLSATCEPADDTLTWTGCGDGGKSVFAKIIGLKPDMFYTCGGTKRQDVLYOYIE
	DPRKNLLIDVPNRCLNYINVALLECYNKRASFSDSKYEPLSYLGFDHVHLVFANVLIPD
	YLKTSRDRIKLWNI"
CDS	314 ..775 /note="ORF 2" /codon_start=1 /protein_id="AAA42895.1" /db_xref="GI:323308" /translation="MTGFTWSRVVPTNRIETTVRNGCFSTSERLVLESKGHDWPN DLLNLMSAKTQADEDALYTELRWNGDGPKIRSHFHITIGSLKCCLRSEORT TAQSGVADEPFETGSPCLNISDDSPGTGTHVVEPERTYCTSTRTQNEI" complement(422..568) /note="ORF 6" /codon_start=1 /protein_id="AAA42896.1" /db_xref="GI:323309" /translation="MEMGTDFORDILSIPPKLVRQRIFGIRGLPGGVHQVPPQIVGP IVAP"
CDS	639..797 /note="ORF 3" /codon_start=1 /protein_id="AAA42897.1" /db_xref="GI:323310" /translation="MTRRRRREVRVCQISRTQARLVLHMWNQKGRIVEVHRGPKTK FNPRCTQV" complement(823..987) /note="ORF 5" /codon_start=1 /protein_id="AAA42898.1" /db_xref="GI:323311" /translation="MTHTLNIPQFYFVPADFQIIIRDIGKEYMHMVEPKITKGFFVF RTECPVLNTF" 1098..1286 /note="ORF 4" /codon_start=1 /protein_id="AAA42899.1" /db_xref="GI:323312" /translation="MNRYVMGGPTIKDSIWFTNLLLCLQCTQTLPSTPIOVSSILLEKK AASLYLPSTCFCAIGRLS"
CDS	BASE COUNT 336 a 323 c 332 g 300 t ORIGIN
	Query Match 100.0%; Score 381; DB 58; Length 1291; Best Local Similarity 100.0%; Pred.No. 2e-99; Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 cagcgacgacgcacaatcctctgdatatcggaacgacgagagacggaggagtccg 60
Db	611 CAGCGGACGATCGCACAAATCCTCTGGNATATCGGACGACGAGAGCGGAAGTCG 670
QY	61 tgtttgccaaatatctoggactcaagcccgaactggttacctacatgtggtggaaccagaa 120
Db	671 TGTTTGCCAAATATCTGGACTCAAGCCCAGCTGGTTTACACATGTGTGGAACAGAA 730
QY	121 aggaacgatgtaccagatacatgagaccaccaaaaagaattaatcctcgatgtaacca 180
Db	731 AGGACGCTATTGTACAGPATCATCGAGACCACCAAAACGAAATTTAATCCTCGATGATCCA 790
QY	181 ggtgtaatttagagtatttaaattatgccctgttgataatggtttaagaacaggcca 240
Db	791 GGTTGTAATTTAGAGTATTTAAATTTATGCCCTGTTAGAATGTGTTAAGAACAGGCATTCA 850
QY	241 gtccggacaaatacgaaccccttagtttatcttgggttcaccactgcatgtactcgat 300
Db	851 GTTCGGACAAATACGAACCCCTTAGTTATCTTGGGTTCGACCATGTGCATGTACTCGTAT 910
QY	301 ttgccaaatgctcgcctgattatttgaaatacacgaggagacagaataaaaaactgtggaata 360
Db	911 TTGCCAATGCTCGCTGATTTTGRAAATCAGCAGGGACAGAAATAAAAACTGTGGAATA 970
QY	361 tttaaagtatgtgtcatctaa 381

[illegible]


```

/protein_id="AA51429.1"
/db_xref="GI:520795"
/translation="WGLEPNKRFNYNDFLHMLACSYRVGSTCFLIACGGSTLS
QQLIFLUSLEILQ"
347 a      231 c      244 g      274 t
BASE COUNT
ORIGIN

```

Query Match 29.5%; Score 112.4; DB 58; Length 1096;
Best Local Similarity 57.9%; Pred. No. 7.2e-22;
Matches 219; Conservative 0; Mismatches 156; Indels 3; Gaps 1;

Qy	4	cggacgacgcgaacatcctctcgatcgacgcgagcggagagcaggaagcccggt	63
Db	564	CTGATGTCGAAGTATCATCTGGGTACGGTCCACGGAGCGCAGGAAGTCAACCT	623
Qy	64	ttgccaaatatctcggactcaagcccgactggttctacacatggttggaaccgagaag	123
Db	624	TGCGAAGATATCTATPACATTAACACCGGATGGGATATATCAACGGTGGAAAGACGTCGG	683
Qy	124	acgtattgtaccagtcacatcgcagaccacaaacgaaatttaactcctcgatgtaccocaggt	183
Db	684	ATATGATGCCATCATACAGATGGATCTGTATAATCATATGGATTATTTGATATCCCCAGAA	743
Qy	184	gtaattagagtatttaaattatgcctgtttagaatgtgttaagacacagcgatcagtt	243
Db	744	GTCATTCAGATTATCTGAAATTTATGCGGTTATAGACACAAATTAGAATAGAGTTTATAAA	803
Qy	244	cggacaaatagcaacc---ccctagtatcttgggttcgaccattgcctagtactogtat	300
Db	804	ATACAAAATAGCAACCATGTGTGATTAGAAAAGATGACAAAATGTCCATGTAATTGTTA	863
Qy	301	ttgccaatgtcctgcctgattatttgaataatcagcagggcagacaataaaactctggaata	360
Db	864	TGGCAATGTGTTGCCTCGATTATTTGTAATAATTTCAAGAGATAGATAAAAAATAATT	923
Qy	361	tttaaagtatgtgtcatc	378
Db	924	GTTCAGAAAGGAAACTTC	941

```

RESULT      6
AR063453
LOCUS       AR063453          1091 bp    DNA
DEFINITION  Sequence 5 from patent US 5846705 .
ACCESSION   AF063453
VERSION     AR063453.1 GI:5992761
KEYWORDS    .
SOURCE      Unknown.
            Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1091)
AUTHORS    Wu,R., You,L. and Soong,T.
TITLE       Nucleotide sequence of two circular SSDNA associated with banana
            bunchy top virus and method for detection of banana bunchy top
            virus
JOURNAL     Patent: US 5846705-A 5 08-DEC-1998;
FEATURES    Location/Qualifiers
             ..1091
             /organism="unknown"
BASE COUNT  343 a      224 c      248 q      276 t

```

	Query Match	29.1%	Score 110.8;	DB 9;	Length 1091;
	Best Local Similarity	57.7%;	Pred No. 2.1e-21;		
	Matches 218;	Conservative	0;	Mismatches 157;	Indels 3; Gaps 1;
Qy	4	cggaacatcgacacatcctctggtatctggagcagacgaggagagacgggaagtcctgt	63		
Db	559	CTCATCATCGAAGTATCATCTGGGTATACGGTCTCTCCGGAGGCCGAAGAAGTCAACCT	618		
Qy	64	tigccaatatctcggactcaagcccgactggtttctacatgtggttggaaccgaaagg	123		

Db	619	TCCGACAGATATCATCAATTAACACCTGGATGGGGATATATCAACCGTGGAAAGACGTCGG	678
QY	124	acgtattgtaccagtagcatcagagaccacgaacgaatttaacctcgcgatgcaccagct	183
Db	679	ATATGATGCCACATCAATAACGATGGATCCTGTATAATCATTTGGATTATGATATCCCGAGAA	738
QY	184	gtaatttagaagtattaaattatgcctgttagaattggttaagaacacagggcattcagtt	243
Db	739	GTCATTTCAGATTATCTGAATTATGGCGTTATAGACAAATTAAGAATAGAGTTTAAATAA	798
QY	244	cggacaaatcacgaacc--ccttagttatcttgggttcgaccatgtcgtgtaclogtat	300
Db	799	ATACAAAATACGAACCATGTGTGATTAGAAAAGATGGCAAAATGTCCAATGTAATTTGTTA	858
QY	301	tgcgaattgctcgtcgtgattatttgaataatcagcaggagacagaataaaactgtggaata	360
Db	859	TGCAAAATGTGTGCTGATTATTTGTAATTCAGAGATAGATAAAAAATAATTAATT	918
QY	361	tttaaagtatgtgtcatc	378
Db	919	GTTGAGAAAGGAACATTC	936
RESULT	7		
AF216222			
LOCUS			
DEFINITION	AF216222	1095 bp DNA VRL	08-MAR-2001
		Banana bunchy top virus satellite S2 replication initiation protein	
ACCESSION	AF216222		
VERSION			
KEYWORDS			
SOURCE	AF216222.1	GI:12004327	
ORGANISM		Banana bunchy top virus.	
REFERENCE		Banana bunchy top virus	
AUTHORS		Viruses; ssDNA viruses; Circoviridae.	
TITLE		1 (bases 1 to 1095)	
		Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.	
		Additional rep-encoding DNAs associated with banana bunchy top	
		virus	
JOURNAL	Arch. Virol.	146 (1), 71-86 (2001)	
MEDLINE	21163975		
REFERENCE		2 (bases 1 to 1095)	
AUTHORS		Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.	
TITLE		Direct Submission	
JOURNAL	Submitted (14-DEC-1999)	School of Life Sciences, Queensland	
		University of Technology, GPO Box 2434, Brisbane, QLD 4001,	
		Australia	
FEATURES		Location/Qualifiers	
source		1..1095	
		/organism="Banana bunchy top virus"	
		/isolate="Taiwan"	
		/db_xref="taxon:12585"	
		/note="satellite S2"	
		64..921	
mRNA		/gene="ORF V1"	
gene		64..921	
		/gene="ORF V1"	
		64..921	
CDS		/gene="ORF V1"	
		/codon_start=1	
		/product="replication initiation protein"	
		/protein_id="AAG44004.1"	
		/db_xref="GI:12004328"	
		/translation="MSSPFWCHTLYNYSAAAREDFELALLKEELNAYVVGDEAVPSTG	
		RKHLOGLSLKSKIKLGLLKKYSKAWHERAGTEQNRGYCSKETLVLEIGTPVP	
		GSKKKLLRFREFPELKMEDPSKYRCLAVESLNARKNSEWHEHLEWQNKLIQH	
		IGVDPDRSLIWIYGPNGEGEKSTFARYLSLKPWGVIYNGKTSMDMHIITMDPNHW	
		IEDIRPSHDLYNGVIEQIKNRVLINTKYPCVIRKQCNVHVIVMANVLPDYCKIS	
		EDIIRKIING"	
BASE COUNT	347 a	219 c	252 g
ORIGIN			277 t

	Query Match	29.0%; Score 110.6; DB 58; Length 1095;		
	Best Local Similarity	58.4%; Pred. No. 2.4e-21;		
	Matches 213; Conservative	0; Mismatches 149; Indels	3; Gaps 1;	
y	4	cggacgatcgcaaatcctctgatatgcggacgagacgaggagacggaagtccgtgt	63	
b	557	CTGATGTCGAAGTATCATCTGGGTATACGGTCCCAACCGGCGAAGGAAAAGTCAACCT	616	
y	64	ttgccaaatattctcgactaacgcccgaactggttctacacatggttgtaaccagaag	123	
b	617	TGCAGATATATCATATAAAACCTGGATGGGGATATATCAACGGTGGAAGACGTCGG	676	
y	124	acgtattgtaccagtcacatcgagaccaccaaaacgaattttaacctctcgatgtaccaggt	183	
b	677	ATATGTCACATCAATCAACGATCGATCCTGATTAATCATTTGGATTATTGATATCCCAGAA	736	
y	184	gtaatttagagttattaaattatgccctgttagaaagtgttaagaacagaggaattcagtt	243	
b	737	GTCATTTCAGATTATCTGAATTAATGGGTTATAGAACAATAAAGATAGAGTTTTTAATA	796	
y	244	cggaacaatacgaacc---cctlagttatcttggttgcaccatgttgcatgtactcgtat	300	
b	797	ATACAAATACGACCAATGTGTGATTAGAAAAAGATCGACAAATATGCCATGTAATTGTTA	856	
y	301	tigccaatgtctgcctgattatttgaataatcagcggggacagaaataaaaactgtggaata	360	
b	857	TGGCAAATGTGTCCTGATTTATGTTAAATTTTTCAGAAAGATAGAAATAAAAATAATTAAT	916	
y	361	tttaa 365		
b	917	GTGA 921		
	RESULT 8			
	BBU12586	1111 bp DNA circular VRL	01-FEB-1995	
	LOCUS	Banana bunchy top virus DNA III ORF VI and ORF CI genes, complete cds.		
	DEFINITION			
	ACCSSION	U12586		
	VERSION	U12586.1 GI:642390		
	KEYWORDS			
	SOURCE	Banana bunchy top virus.		
	ORGANISM	Banana bunchy top virus.		
	REFERENCE	Viruses; ssDNA viruses; Circoviridae.		
	AUTHORS	1 (bases 1 to 1111)		
	JOURNAL	Wu,R.-Y. and You,L.-R.		
	TITLE	Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related viruses		
	UNPUBLISHED			
	2 (bases 1 to 1111)			
	Direct Submission			
	Submitted (25-JUL-1994)	Rey-Yuh Wu, Agricultural Biotechnology Division, Development Center for Biotechnology, 81 Chang Hsing Street, Taipei, 10671, Taiwan, ROC		
	Location/Qualifiers			
	1..1111	/organism="Banana bunchy top virus"		
		/db_xref="taxon:12585"		
		/clone="H-4"		
		1..8		
		15..49		
		79..933		
		/note="ORF V1"		
		/codon_start=1		
		/protein_id="AAA61875.1"		
		/db_xref="GI:642391"		
		/translation="MSSFKWCFTLNYSSAAREDFLALLKEELNAYVGVDEVAPSSG QKHGLYLSLKKS IKLGKKKYSRAHWERARGSDENAKYCKETLLELGPASQ OGNRKLSMYRSRSPERIKIPEPIHYRTSVKLKKPKFEFVHPCLDRWQIQLTEA IDEEPDDDTIFWVGPNCEKGSTYAKSLMKDWFTYTRGGKNILFSYVDGESEKHI EEDPCDCHVTGPNCEALDNPTSEKYEKIIVNTHTVMANFMPEFCCKIE		
	TATA_signal			
	stem_loop			
	CDS			
	CDs			
	complement(497..643)			
	/codon_start=1			
	/product="ORF C1"			
	/protein_id="AAA61876.1"			
	/db_xref="GI:642392"			
	/translation="MTSHMLISPHYHSDHRPRSFDLHVLPQLPPSIESAMVYRGKDE RIPP"			
	BASE COUNT	344 a 213 c 260 g 294 t		
	ORIGIN			
	Query Match	25.2%; Score 96; DB 58; Length 1111;		
	Best Local Similarity	57.6%; Pred. No. 3.9e-17;		
	Matches 212; Conservative	0; Mismatches 150; Indels	6; Gaps 2;	
	QY	2	acggacgatcgcaaatcctctgatatgcggacgagacgaggagacggaagtccgt 61	
	DB	570	ACCAGATGATCGAACGATCTTCTGGGTCTATGTCGGAATGTAATGAGGGAATCAAC 629	
	QY	62	gtttgccaaatattctcgactcaagcccgaactggttctacacatggttgtaaccaga 121	
	DB	630	ATATGCGAAGTCATTT---AATGAAGAAGACTGGTCTTACACCAGAGTGGGAAGAAGA 686	
	QY	122	ggacgtattgtaccagtcacatcgagagccccaaaaacgaa---attaatcctcgatgtacc 178	
	DB	687	GAACATAGCTGTTCTCTTAGCTGGACGAAGGATCTGAGAAGCATATTGTTATTGATATTC 746	
	QY	179	caggtgttaattagagtatattaaattatgccctgttagaaagtgttaagaacagggcatt 238	
	DB	747	TCGCTGTAATCAGGATTATTTAAATATATGATCTTATAGGAGCATTAAGGATAGGTGAT 806	
	QY	239	cagttcggacaatacgaaccccttagttatcttggtttcgaccatgtgcattgactcgt 298	
	DB	807	AGAGTACTAAATATAAACCTATTAAAGTTAGTTGAATTAATATACATGTAATTGT 866	
	QY	299	atttgccaatgtcctcgctgattatttgaataatcagcggggacagaaataaaaactgtgga 358	
	DB	867	AATGCGTAATTTTCATGCCCAAGATCTCTGTAATAATCTCCGAAGATAGATAAAGATTATTTA 926	
	QY	359	tatttaaa 366	
	DB	927	TTGTTAAA 934	
	RESULT 9			
	FBNVVREP	1002 bp DNA	VRL	31-JAN-1995
	LOCUS	FBNVV gene for replicase.		
	DEFINITION			
	ACCSSION	X80879.1	GI:643114	
	VERSION	X80879.1		
	KEYWORDS	replicase.		
	SOURCE	Viruses.		
	ORGANISM	Unclassified.		
	REFERENCE	1 (bases 1 to 1002)		
	AUTHORS	Katul,L., Maiss,E. and Vetten,H.J.		
	TITLE	Sequence analysis of a faba bean necrotic yellows virus DNA component containing a putative replicase gene		
	JOURNAL	J. Gen. Virol. 76 (Pt 2), 475-479 (1995)		
	MEDLINE	95146992		
	REFERENCE	2 (bases 1 to 1002)		
	AUTHORS	Katul,L.		
	TITLE	Direct Submission		
	JOURNAL	Submitted (03-AUG-1994) L. Katul, Biologische Bundesanstalt, fuer Land und Fortwirtschaft, Inst f Biochemie und Pflanz., Messeweg 11-12, 38104 Braunschweig, FRG		
	LOCATION/Qualifiers			
	source	1..1002		
		/organism="Viruses"		
		/isolate="SSV 292-88"		

Query Match 24.4%; Score 93; DB 59; Length 1004;
 Best local Similarity 55.3%; Pred. NO. 2.9e-16;
 Matches 202; Conservative 0; Mismatches 160; Indels 3; Caps 1;

Qy	2	acgcgacgatcgacaatcctctggatatgcgcgacgagcaggagagacggaagtccgt	61
Db	551	ACCACAGCAGCAGAAGATACTCGGTTTACGGATCTGATGGAGGAGAGGTAAATCAAC	610
Qy	62	gtttgccaaatatctcgactcaagcccgcactggttctcaccatgtggtggaaaccagaaa	121
Db	611	CTTCGCGAAGGAATT--AAITTAAGTATGGATGTTTTATCTGCAGGAGAAAACGCCA	667
Qy	122	ggcgttatgtaccagtcatactcgaggaccocaaaaacgaatttaactoctcgatgiaccgag	181
Db	668	GGATAATCTGTATATGTTATGCTCAAGACCCAGACAGAAAAATATCGCTTGATGTTCCCAG	727
Qy	182	gtgcaatttagagtatttaaatattgccctgttgagaatggttaaagaacaggcattccag	241
Db	728	GTGTTCTTCGGAGATGATGAAGCTAACGCTATCGAGATGTTGNAGACAGAGTAATTGC	787
Qy	242	ttcgacaaaatacgaaccccttagttatcttgggtctcgacctgtgcatgtactcgtatt	301
Db	788	AAGTACAANAATATAGGCCCGTAGATCTTTGTTTGAAGAAAAGTTTCATTTAATTGTGT	847
Qy	302	tgccaatgtctgcctgattatttggaaatacagcagggcacagaataaaaactgtggaaatal	361
Db	848	TGCAAAAGTGTCACTCCCTGACCCACAAAAATAAGTCAGCAGAAATTGTAATTATCAATTG	907
Qy	362	ttaaa	366
Db	908	TTGAA	912

RESULT 12	AB0000920	30-OCT-1998
LOCUS	AB0000920	
DEFINITION	Milk vetch dwarf virus genome segment 1 encoding viral replication-associated protein, complete sequence.	
ACCESSION	AB000920	
VERSION	AB000920.1	GI:3808179
KEYWORDS	viral replication-associated protein.	
SOURCE	milk vetch dwarf virus (lab_host:Pisum sativum) DNA.	
ORGANISM	milk vetch dwarf virus	
REFERENCE	Viruses; ssDNA viruses; Nanovirus.	
	1 (bases 1 to 1007)	

Direct Submission
Submitted (08-FEB-1997) to the DDBJ/EMBL/GenBank databases.
Yoshitaka Sano, Kyoto Institute of Technology, Department of
Applied Biology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan
(Tel:075-724-7764, Fax:075-724-7764)

2 (sites)
Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. and Kojima, M.
Sequences of ten circular ssDNA components associated with the mill
vetch dwarf virus genome
J. Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
99094638
On Oct 29, 1998 this sequence version replaced gi:3798651.
Sequence updated (30-Sep-1997).

```

FEATURES
    source
        Location/Qualifiers
            1..1007
                /organism="milk vetch dwarf virus"
                /virion
                /db_xref="taxon:67585"
                /chromosome="segment 1"
                /lab_host="Pisum sativum"
            1..33
                stem_loop
            51..57
                TATA_signal
            91..936
                CDS
                    /function="helicase"
                    /note="putative"
                    /codon_start=1
                    /product="viral replication-associated protein"

```

179 cagqgtgaatttagagatatttaaatatgcocgtttagaatgtgttaagaacagagggcatt 238

Db	747	TCCGCTGTAATCAGGAATTATTTAAATATATGATGTTATAGAGCAATTAAAGGATAGGGTGAT	806
Qy	239	cagttcggaacaatacgaaccccttagttatcttgggttcgaccatgtgcaigtactcgt	298
Db	807	AGAGACTACTAATAATAAACCTATTAAAGTTAGTTGAATTGAATATATACATGTAATTTGT	866
Qy	299	atttgccaatgtcctgcctgattatttgaataacagcagggcacagaataaaaactgtgaa	358
Db	867	CATGGGTAAATTCATGCCAGAAATCTTGTAATAATCTCCGAAGATAGATAAAGATTATTTA	926
Qy	359	tatttaaa 366	
Db	927	TTGTTAAA 934	
RESULT 11			
NV5968			
LOCUS	NV5968	1004 bp	DNA circular VRL
DEFINITION	faba bean necrotic yellows virus Cl-Eg gene.		
ACCESSION	AJ005968		
VERSION	AJ005968.1 GI:3550536		
KEYWORDS	Cl-Eg gene; component 1-Eg; putative; rep protein; replication associated protein.		
SOURCE	faba bean necrotic yellows virus.		
ORGANISM	faba bean necrotic yellows virus		
REFERENCE	1 (bases 1 to 1004)		

FILE	biochem	Submitted (06-MAY-1998)	Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
REFERENCE	2	(bases 1 to 1004)	
AUTHORS	Katul,L., Timchenko,T., Gronenborn,B. and Vetter,H.J.		
TITLE	ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome		
JOURNAL	J. Gen. Virol.	79 (Pt 12), 3101-3109	(1998)
MEDLINE	99094637		
FEATURES	Location/Qualifiers		
source	1..1004		

```

stem_loop
gene
1..31
63..1003
/gene="Cl-Eg"
63..911
/gene="Cl-Eg"
/function="putative replication associated (rep) protein"
/codon_start=1
/product="component 1-Eg"
/protein_id="CAA06791.1"
/db_xref="GI:3550537"

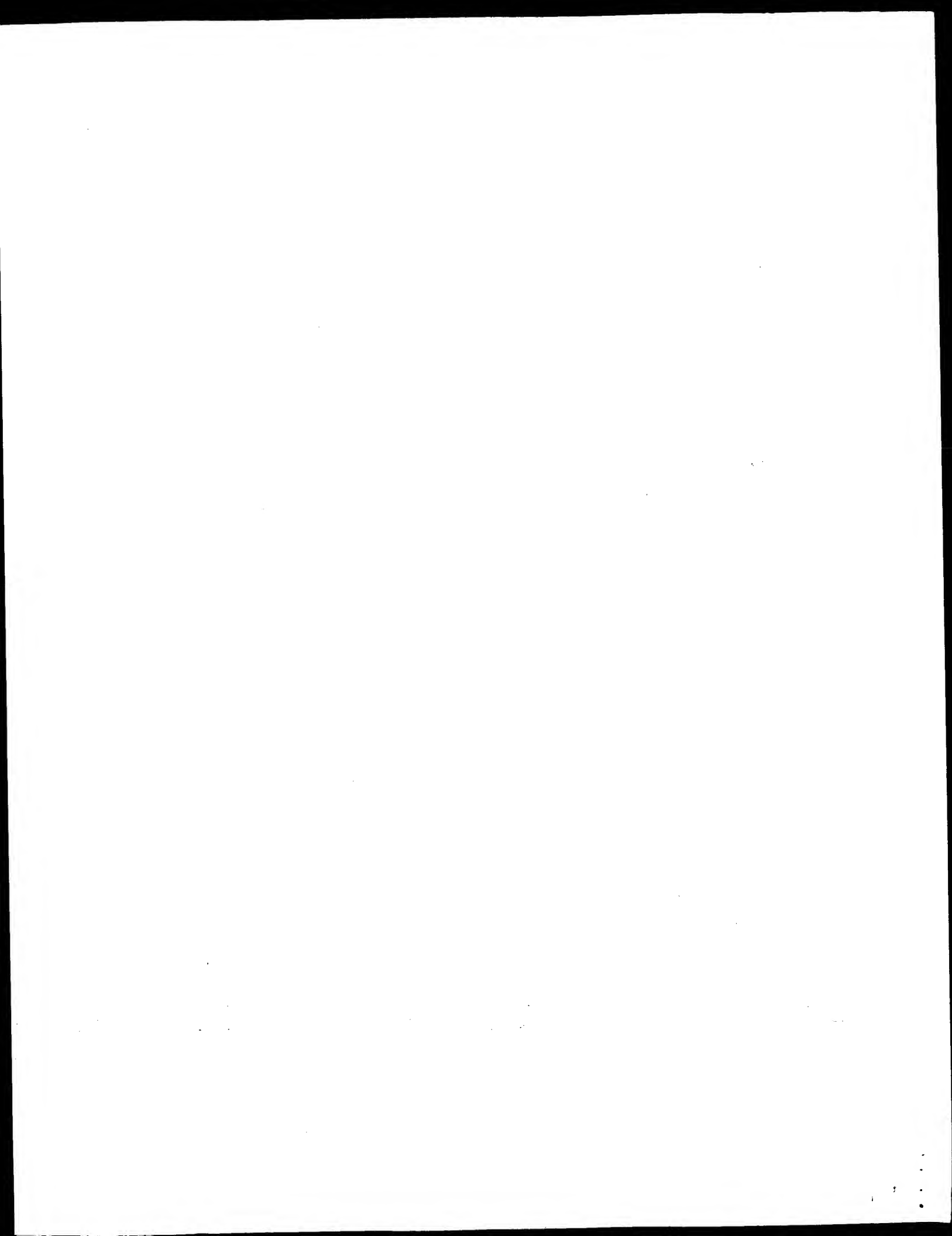
```

polyA_signal	911..916	/gene="Cl-Eg"			
TATA_signal	938..1003	/gene="Cl-Eg"	224 g	296 t	
BASE COUNT	305 a	179 c	224 g	296 t	
ORIGIN					

[illegible]

Db 542 ACCAGATGACCGTCAAAATTATTGGGTGTATGGCCACATCGAGGAGAGGTAAGCGC 601
QY 62 gtttgccaaatatctcggaactcaagcccgactcgttctacacatgtgtggaaccagaaa 121
Db 602 GTTTCGGAATACITTAACCTTGAAGGAGGTTGGTGTATACAGCTGGAGGGAAGCGAC 661
QY 122 ggaacgtattgtaccagtaacatcgaggaccacaaacgaaatttaactcctcgatgtacccag 181
Db 662 AGATATGCTTTACAGCTATTCAITGGACCCACATGCCAGTATGCATTGCATATCCCGAG 721
QY 182 gtgt---aatttagagtatttaaatatgcccgttagaagtgttaagaacagggcatt 238
Db 722 ATGTACAGAGAGAATAATATTAATTCTGTAAATGAACAGATTAAAAATAGAGTTAT 781
QY 239 cagttcggacaaatacgaaccccttagttatcttgggttcg---accatgtgcattgtact 295
Db 782 TATTAATACAAAATATGAACCTGTACTATTAGAGACGATGGTCATAATGTTTCATGTAAT 841
QY 296 cgtatttgccaaatgctcgtcgtattatttgaataatcagcaggggacagaataaaaactgtg 355
Db 842 TGTATTTTGTAAATTTTACCTGTGTAAACAGAAATCAGTGAAGATAGATAAAAAATAAT 901
QY 356 gaattattaaagt 368
Db 902 AAATGTTGAAAT 914

Search completed: October 24, 2001, 11:42:48
Job time: 6353 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:44 ; Search time 5701.85 seconds
(without alignments)
1581.539 Million cell updates/sec

Title: US-09-462-955-1-copy_409_991
Perfect score: 583
Sequence: 1 actgtctgtgagtcataag.....ttaaagtatgtgtcatctaa 583

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: gb_ba1:
17: gb_ba2:
18: gb_fun:
19: gb_htgo_hum:
20: gb_htgo_inv:
21: gb_htgo_rod:
22: gb_htg_hum1:
23: gb_htg_hum2:
24: gb_htg_hum3:
25: gb_htg_hum4:
26: gb_htg_hum5:
27: gb_htg_hum6:
28: gb_htg_hum7:
29: gb_htg_hum8:
30: gb_htg_inv1:
31: gb_htg_inv2:
32: gb_htg_other:
33: gb_htg_rod:
34: gb_hum1:
35: gb_hum2:
36: gb_hum3:
37: gb_hum4:
38: gb_hum5:
39: gb_hum6:
40: gb_hum7:
41: gb_in:
42: gb_om:
43: gb_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_ro:
49: em_rod:
50: em_sy:
51: em_un:
52: em_vi:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_vil:
59: gb_vil2:
60: gb_vil3:
61: gb_vil4:
62: gb_vil5:
63: gb_vil6:
64: gb_vil7:
65: gb_vil8:
66: gb_vil9:
67: gb_vil10:
68: gb_vil11:
69: gb_vil12:
70: gb_vil13:
71: gb_vil14:
72: gb_vil15:
73: gb_vil16:
74: gb_vil17:
75: gb_vil18:
76: gb_vil19:
77: gb_vil20:
78: gb_vil21:
79: gb_vil22:
80: gb_vil23:
81: gb_vil24:
82: gb_vil25:
83: gb_vil26:
84: gb_vil27:
85: gb_vil28:
86: gb_vil29:
87: gb_vil30:
88: gb_vil31:
89: gb_vil32:
90: gb_vil33:
91: gb_vil34:
92: gb_vil35:
93: gb_vil36:
94: gb_vil37:
95: gb_vil38:
96: gb_vil39:
97: gb_vil40:
98: gb_vil41:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	583	100.0	1291	58	CFDCG
2	132	22.6	1004	59	M29963 Coconut fol
3	132	22.6	1007	58	AJ005966 faba bean
4	116.4	20.0	1096	9	AJ132187 faba bean
5	116.4	20.0	1096	58	AR063452 Sequence
6	114.8	19.7	1091	9	L32167 Banana bunc
7	114.4	19.6	1095	58	AR063453 Sequence
8	108.6	18.6	1111	58	AF216222 Banana bu
					U12586 Banana bunc

9 107 18.4 1127 58 BB012587
10 104 17.8 1015 58 FBE132185
11 103.8 17.8 1109 58 AF216221
12 103 17.7 1022 58 AB009047
13 102.2 17.5 1106 9 AR063451
14 102.2 17.5 1106 58 BYTV1
15 101.6 17.4 1014 59 NYV5964
16 99.2 17.0 1017 59 SCU16735
17 98.8 16.9 1009 58 AB000921
18 98 16.8 1095 59 U023112
19 96.4 16.5 1002 58 FBNVREP
20 96 16.5 1000 58 AB000922
21 94.4 16.2 1004 59 NYV5968
22 94.4 16.2 1022 59 SCU16731
23 92.8 15.9 1007 58 AB000920
24 91.4 15.7 1375 58 CLE132345
25 88.2 15.1 1367 58 AYE238493
26 86.6 14.9 1376 58 CLE132344
27 82.8 10.8 1110 9 AR010230
28 59.6 10.2 982 9 AR112007
29 59.6 10.2 1110 9 AR010227
30 59.6 10.2 1110 58 BB018077
31 59.6 10.2 1111 9 AR010225
32 59.6 10.2 1111 9 AR010228
33 59.6 10.2 1111 9 AR010240
34 59.6 10.2 1111 58 AF102780
35 59.6 10.2 1111 59 S56276
36 58.6 10.1 1103 9 AR010232
37 58.6 10.1 1105 9 AR010229
38 58 9.9 1109 9 AR010231
39 58 9.9 1111 58 AF238874
40 57 9.8 1103 59 U97525
41 57 9.8 1104 9 AR010233
42 57 9.8 1111 9 AR010226
43 56.4 9.7 1111 58 AF238875
44 53.8 9.2 1103 58 AF246123
45 52.2 9.0 1103 58 AF246123

ALIGNMENTS

RESULT 1
CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993
LOCUS Coconut foliar decay virus, complete genome.
DEFINITION M29963
ACCESSION M29963.1 GI:323306
VERSION circular; complete genome.
KEYWORDS Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
SOURCE Coconut foliar decay virus
ORGANISM Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1291)
AUTHORS Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
JOURNAL Virology 176, 648-651 (1990)
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by W.Rhode, 15-MAR-1989, for release after publication.
FEATURES
source Location/Qualifiers
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40..70
/note="stem-loop structure"
103..975
/notes="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSIRRWCFILNYETEEFAANVVRRTESLNLVYIVGVDEVPAS
TGORHLQGFTHLKTGRRLQGLTVLGNDRHLHLEPTRGSDQNRDYCKSERVLLHGVGP

TRGVKRPRLAORFAEPEDELRLDPGGYRRCVVGASVETWRAAENPPFFPVHNO
LEVLISAISPADRRITILICGRDGGGKSVFAKYLGLKPMWFYTCGGTRKDVLYQYIE
DKNALILDVPCNLELYLALLECCKNRAFSDDKYEPVLSYLGFDHVLVFAVNLDP
YKLISDRILKLNWI
314..775
/notes="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MIGFTWSPVPTNRIETVVRNGCFSTSRVLVLESGHWDPN
DLRLNLSNAWKTQADIEDALYLRWQDGLPKIRSHFHTIGSLKCLRSRQRT
IAOSSYADETEETGSPCLNPSDSPTGSTHVEPERTYCTSTRTONEI"
complement(422..568)
/notes="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGYDFORPILSIPPKLRVQIRLGLPGGVHVPQIVGP
IVAF"
639..797
/notes="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MTRRRRREVRCQISRTQARLVLMWVWVQGRVVPVHRGPKTK
FNPRCTGV"
complement(823..987)
/notes="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHLNIPOFYSPADFQIIRODIDIGKEYEMHVEPKITRGFVVF
RTECPVLNPF"
1098..1286
/notes="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGPGPTIKDSIWRNLLLCLOQTPLSTPQVSSLLEKK
AASLYLPSICFCAIGRLS"
BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 100.0%; Score 583; DB 58; Length 1291;
Best Local Similarity 100.0%; Pred. No. 2.3e-161; Indels 0; Gaps 0;
Matches 583; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 actcgtcctggagtcacaaagccacgattgcccacgatttgcctgaggaacctgatgaa 60
Db 409 ACTCGTCCTGGAGTCNAAGGCCAGATTGCCCAACGATTGCTGAGGAACCTGATGAA 468
QY 61 ctccgctggaagaccagcgggataccgaagatgcgtgttacacggagcttcggtggaa 120
Db 469 CTCGGCTGGAAGACCCAGCGGATACCGAGATCGTGTGACACGGAGCTTCGGTGGAA 528
QY 121 tggacaagatggcgcgtgaaatccgttccattccattccatcacattcgacacttgaa 180
Db 529 TGGACAAGATGGGCGCTGAAATCCGTTCCCTTTTCCATATCACAATTGGCAGCTTGA 588
QY 181 gtgctgtctgcgacgcgagcgcgacgacgacgcgacgcgacgcgacgcgacgcgacg 240
Db 589 GTGCTGTCTGCGATCGGAGCGGACGAGGATCGCACAACTCTCTGGATATGCGGAGA 648
QY 241 gacggaggaacggggaagtcctgtttgccaataatctcggactcaacggactggttc 300
Db 649 GACGGAGGAGACGGGAAGTCCTGTTTGGCAATATCTCGGACTCAAGCCGACGGTTC 708
QY 301 tacacatgtgtgaaacgaagacgctatgtaccagatcatcatcagagaccgaacga 360
Db 709 TACACATGTGTGGAACCAAGAGAGAGAGATGTGTACCATCATCATGAGAGACCAACGA 768
QY 361 aatttaattcctcgatgtaccagaggttaatttagagattttaatttagcctgttagaa 420

```

Db 769 AATTTAAATCCTCGATGTCACCGAGTGTAATTTAGAGTATTTAAATATGCCCCCTGTAGAA 828
QY 421 tctgttaagaacaggaattcagcttcgagcaataacgaaccccttagttatcttggttc 480
Db 829 TGTGTTAAGAACAGGCAATTCAGTTCGACAAATACGAACCCCTTAGTTATCTTGGTTC 888
QY 481 gacctgtgcatgtactgtatttgcgaatgctcgtcgtatatttgaaatacagcag 540
Db 889 GACCATGTGCATGTACTCGTATTTGCCAATGCTCTGCTGCTGATTTTGAANAATCAGCAGG 948
QY 541 gacagaataaaactgtgaatatttaagaatgtgtcatctaa 583
Db 949 GACAGAATAAAACTGTGAATATTTAAAGTATGTGTCTCATCTAA 991

RESULT 2
LOCUS NIV5966 1004 bp DNA circular VRL 13-JAN-1999
DEFINITION faba bean necrotic yellows virus C9 gene.
ACCESSION AJ005966
VERSION AJ005966.1 GI:3550532
KEYWORDS C9 gene; component 9; putative; rep protein; replication associated
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES
source
1. .1004
/organism="faba bean necrotic yellows virus"
/virion
/isolate="SV292-88"
/db_xref="taxon:59817"
73. .918
/gene="C9"
73. .918
/gene="C9"
/function="putative replication associated (rep) protein"
/codon_start=1
/product="component 9"
/protein_id="CAA06789.1"
/db_xref="GI:3550533"
/db_xref="SPTREMBL:Q91252"
/translation="MSAVNNVFTLNFAGEVPLSFDERVOYAVWQHERVNHDIQGV
QLKKAMNIVKNIIGGNPHLEKMGKIEASAYAQKEESRVAGPWSYGLLKKGSH
RKIMELLPDENLEEPQYRRAMWSAMDESKLAEEGFPYMFYSWOETVLGLIE
EPNDRTIIVYGPNGEKGSGFGKFLGKKDLYLPCKTQDMTYMLKPNKANYVMD
IPRCSNYLNYQFMELIKNRTIYSYKPEVPGCIINNIHIVLANVLPDYKISQDRI
KIIYC"
BASE COUNT 352 a 159 c 226 g 267 t
ORIGIN

Query Match 22.6%; Score 132; DB 59; Length 1004;
Best Local Similarity 54.0%; Pred. No. 4.5e-28;
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 68 tgaagaccacagcgatataccagagcttcgtgtacacgagcttcggtggaatggacaa 127
Db 419 TGGGAAGACCCAGAAATACAGAGAGCTATGCTGTCGCCCATGAGCAATCTCGGA 478

```

```

QY 128 gatggcgctgaaataccgttccatttccatatacaaatggcagcttgaagctcgt 187
Db 479 AGCTTGGCGAAGAGAGAGGCTTTCCCTATATGTTTACAGCTGGCAAGAACAGTGTGG 538
QY 188 ctgcgatcgagagccagcgacgacgacgacgacgacgacgacgacgacgacgacg 247
Db 539 GCCTATTAGAAAGAGAGCCCAATGACCGTACTATTATTTCGGTCTACGGCCCAATGGTA 598
QY 248 gagcgggaagctcgtgttgcgaataatctcggaactcaagcccgactggttclacacat 307
Db 599 ATGAAGGAATAATCACAGTTTGGTAAATTCCTGGCATTTAAAAAAGATTACCTTTTATTAC 658
QY 308 gtggtggaaccagagagcgatctgtaccagtcacatcgagcaccacaaacaaatttaa 367
Db 659 CTGGAGGTAAACCCCAAGATATGACATATATGATTAATGAAAAATCCAAAGGCAATGTTG 718
QY 368 tctcgtatgaccaggtgtaattagagttatttaattatccctgttagaattgttta 427
Db 719 TGAATGGATATTCCTCGTGTAAATCTGAATTTAAATTAACCAATTTATGGAATTAATTA 778
QY 428 agaacaggcattcagcttcggacaaatacgaaccccttagttatcttgggttcgaccatg 487
Db 779 ARAATGAACCATATATAGTTATAAATATGAACCACTTGGATGATTAATAATAATAAAA 838
QY 488 tgcattgactcgtatttgcgaatgctcgtcgtgatttattgaaatacagcagggacacaa 547
Db 839 TACATGTAATTTGATTAGCTAATGATTCCCTGATTAAGAAAAATTAGTCAGGATAGAA 898
QY 548 taaaactgtggaatatatttaa 567
Db 899 TTAATAATAATTTATTGTTAA 918

```

RESULT 3

```

FBEL32187
LOCUS faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian EV1-93.
DEFINITION
ACCESSION AJ132187
VERSION AJ132187.1 GI:4995171
KEYWORDS C9-Eg gene; rep protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus
REFERENCE 1 (bases 1 to 1007)
AUTHORS Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
99094637
JOURNAL
MEDLINE
REFERENCE 2 (bases 1 to 1007)
AUTHORS Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetten,H.J. and Gronenborn,B.
TITLE A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
J. Virol. 73 (12), 10173-10182 (1999)
20027244
REFERENCE 3 (bases 1 to 1007)
AUTHORS Katul,L.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. biol. Sicherheit, Biologische Bundesanstalt f. Land- u. Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
FEATURES
source
1. .1007
/organism="faba bean necrotic yellows virus"
/virion
/isolate="Egyptian EV1-93"
/db_xref="taxon:59817"
/lab_host="Vicia faba"

```

Thu Oct 25 13:08:07 2001

us-09-462-955-1_copy_409_991.rge

```

stem_loop      1. .39
gene           73. .1003
              /gene="C9-Eg"
CDS            73. .918
              /gene="C9-Eg"
              /function="replication-associated protein"
              /note="32.53 kDa"
              /codon_start=1
              /evidence=experimental
              /product="Rep protein"
              /protein_id="CAB4027.1"
              /db_xref="GI:4995172"
              /db_xref="SPTREMBL:Q9WIK2"
              /translation="MSAVNNVFTLNFAGVPLSDRVOYAVWQHERVNHHDHQTGVI
              QLKMKMTVKNIIGCNPHLEKMGSIIEASAYAKESRVAGPWSGELLKGSBK
              RKLMLKDPENLEEPQYRRAMASAMDESRLAEKGGFFTYLSWQETVLGLLEE
              EPNDRIITWYGPNGNEGKSGFGKFLGLKKDLYLPGGKTQDMTLMKMPKANYMD
              IPRCNSYLNQFMELIKNRTIPSYKPEVPGCIINNKIHVILANVLDPYEKISQDRI
              KIIYC"
TATA_signal    939. .944
              /gene="C9-Eg"
polya_signal   998. .1003
              /gene="C9-Eg"
BASE COUNT    353 a 161 c 227 g 266 t
ORIGIN
1
22.6%; Score 132; DB 58; Length 1007;
Best Local Similarity 54.0%; Pred. No. 4.5e-28;
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
QY 68 tggagacccagcggtacccagagatcggtgttacacggagcttcgtggaatggcaa 127
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 TGGAGAGACCCAGAAATACAGAGAGCGATGGCTTGGTCGCCATGACGAATCTCGGA 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 gatggccctgaaatccgttccattccatatacacaattggcagcttgaagtgcgt 187
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 AGCTTCTGAAGAGGAGCGCTTCCCTATACGCTTTACAGCTGGCAGAACAGTGTGG 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 188 ctgcgtagcagagccagcggagatcgacaaatccctctgatcgagcagcagcggag 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 GCCTATTAGAAGAGAGCCCAATGACCGCTATTATTATTGGTCTACGGCCCAATGGTA 598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 gacacggagtcggtttgccaatctcgcgactcaagccgactgggttctacacat 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 ACCAAGGAAATACACAGTTTGGTAATCTCTGGGATTAAAAAGATTACCTTTATTAC 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 gtggtgaaccagaaaggcgtattgtaccaglacatcagagagcccaacagaaatttaa 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 659 CTGGAGGTAAACCCCAAGATATGACATATATGTAATGAAAAATCCAAAGCAATGTTG 718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 tccctcgtgtaccagtgtaatttagaattattaaattatgcctgttagaattgttta 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 719 TGATGGATATCTCGCTTGTGAATTTCTGAATATTAAATTTATCAATTTAGCAATTA 778
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 agaacaggcatcagcttcggacaaatacgaaccccttagttatcttgggttcgaccatg 487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 779 AAAATAGAACCATATTAGTTTAAATATGACACAGTTGGATGATTAATAATAATAAAA 838
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 tgcgtactcgtatttggccaatgctcctgcctgattatttgaatacagcagggacagaa 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 839 TACATGTAATTGTATTAGCTTAATGTATTGCTGATTGAAAAAATTAGTCAGGACAGAA 898
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 taaactgtggaattatttaa 567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 899 TTAATAAATTTATTGTGTTAA 918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
LOCUS AR063452 1096 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 4 from patent US 5846705.
ACCESSION AR063452

stem_loop      1. .39
gene           73. .1003
              /gene="C9-Eg"
CDS            73. .918
              /gene="C9-Eg"
              /function="replication-associated protein"
              /note="32.53 kDa"
              /codon_start=1
              /evidence=experimental
              /product="Rep protein"
              /protein_id="CAB4027.1"
              /db_xref="GI:4995172"
              /db_xref="SPTREMBL:Q9WIK2"
              /translation="MSAVNNVFTLNFAGVPLSDRVOYAVWQHERVNHHDHQTGVI
              QLKMKMTVKNIIGCNPHLEKMGSIIEASAYAKESRVAGPWSGELLKGSBK
              RKLMLKDPENLEEPQYRRAMASAMDESRLAEKGGFFTYLSWQETVLGLLEE
              EPNDRIITWYGPNGNEGKSGFGKFLGLKKDLYLPGGKTQDMTLMKMPKANYMD
              IPRCNSYLNQFMELIKNRTIPSYKPEVPGCIINNKIHVILANVLDPYEKISQDRI
              KIIYC"
TATA_signal    939. .944
              /gene="C9-Eg"
polya_signal   998. .1003
              /gene="C9-Eg"
BASE COUNT    353 a 161 c 227 g 266 t
ORIGIN
1
22.6%; Score 132; DB 58; Length 1007;
Best Local Similarity 54.0%; Pred. No. 4.5e-28;
Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
QY 68 tggagacccagcggtacccagagatcggtgttacacggagcttcgtggaatggcaa 127
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 TGGAGAGACCCAGAAATACAGAGAGCGATGGCTTGGTCGCCATGACGAATCTCGGA 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 gatggccctgaaatccgttccattccatatacacaattggcagcttgaagtgcgt 187
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 AGCTTCTGAAGAGGAGCGCTTCCCTATACGCTTTACAGCTGGCAGAACAGTGTGG 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 188 ctgcgtagcagagccagcggagatcgacaaatccctctgatcgagcagcagcggag 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 GCCTATTAGAAGAGAGCCCAATGACCGCTATTATTATTGGTCTACGGCCCAATGGTA 598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 gacacggagtcggtttgccaatctcgcgactcaagccgactgggttctacacat 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 ACCAAGGAAATACACAGTTTGGTAATCTCTGGGATTAAAAAGATTACCTTTATTAC 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 gtggtgaaccagaaaggcgtattgtaccaglacatcagagagcccaacagaaatttaa 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 659 CTGGAGGTAAACCCCAAGATATGACATATATGTAATGAAAAATCCAAAGCAATGTTG 718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 tccctcgtgtaccagtgtaatttagaattattaaattatgcctgttagaattgttta 427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 719 TGATGGATATCTCGCTTGTGAATTTCTGAATATTAAATTTATCAATTTAGCAATTA 778
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 agaacaggcatcagcttcggacaaatacgaaccccttagttatcttgggttcgaccatg 487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 779 AAAATAGAACCATATTAGTTTAAATATGACACAGTTGGATGATTAATAATAATAAAA 838
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 tgcgtactcgtatttggccaatgctcctgcctgattatttgaatacagcagggacagaa 547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 839 TACATGTAATTGTATTAGCTTAATGTATTGCTGATTGAAAAAATTAGTCAGGACAGAA 898
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 taaactgtggaattatttaa 567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 899 TTAATAAATTTATTGTGTTAA 918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BYTV2 1096 bp DNA VRL 30-OCT-1994
LOCUS Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes,
DEFINITION complete cds.
ACCESSION L32167
VERSION L32167.1 GI:520791
KEYWORDS stem loop.
SOURCE Banana bunchy top virus DNA.
ORGANISM Banana bunchy top virus

VERSION AR063452.1 GI:5992760
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1096)
AUTHORS Wu,R., You,L. and Soong,T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana
bunchy top virus and method for detection of banana bunchy top
virus
JOURNAL Patent: US 5846705-A 4 08-DEC-1998;
FEATURES
    Location/Qualifiers
        1..1096
            /organism="unknown"
BASE COUNT 347 a 231 c 244 g 274 t
ORIGIN
1
20.0%; Score 116.4; DB 9; Length 1096;
Best Local Similarity 52.8%; Pred. No. 1.8e-23;
Matches 305; Conservative 0; Mismatches 261; Indels 12; Gaps 2;
QY 6 tctctggagtcacaaaggccagcagattggcccaacgatttgcaggagaccctgatgaactcgc 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 373 TCCTGGTTCGAAAGAGCGCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 66 cctggaagacccagcggtacccagagatcggttgttacacggagcttcggtggaatggac 125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 433 GATGGACATCCATCCCAAGTATCGCAGATGCTTGGCAGTGGAAATCAATTAAGATGCCAG 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 126 aagatggccgctgaaatccggttccattccatatacacaattggcagcttgaagtgc 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 AATTAATTCGGAATGGGTTCACGAACATA-----AAAGAATGGCAAAATAAATTAAT 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 gtctcgcgtcggagagccagcggagatcgacaaatccctctggatcgcgagcagcagc 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 TCAACACATCGAAGGTCTCTGATGATCGAAGTATCATCTGGGTATACGCTGCCAAGCG 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 246 aggaagacgggagtcggtgttggccaaatatactcggactcaagcccgactgggttctacac 305
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 604 AGCGAAGGAAAGTCAACCTTCGCAAGATATCTATCATTTAAACCCCGGATGGGATATAT 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 306 atgtggtggaacagaaaggcgtattgtaccagtcagagcagcccaaacgaaattt 365
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 664 CAACGCTGGAAGACAGCTCGGATATGATGCACATCATACAGATGGATCCTGATAATCAT 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 366 aatcctcgtatgccagtggttaatttagagttatttaattatccctgttagaattgt 425
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 724 GATTATTGATATCCCGAGAGTCAATTCAGATTATCTGAATTTATGCGGTATAGAACAAAT 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 426 taagaacagggcattcagtcggacaaatacgaaccc--ccttagttatcttgggttcga 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 784 TAAGAATAGAGTTTAAATAAATACAAATACGAACCATGTGTGATTAGAAAGATGGACA 843
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 483 ccatgctcgtactcgtatttggccaatgtcctgcctgattatttgaataatcagcagga 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 844 AATATGCCAATGTAATTTGTTATGCGAAATGTGTGCTGATTATTGTTAAATTCAGAAGA 903
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 543 cagaataaaactgtggaatatttaaatgattgtgtcatc 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 904 TAGAATAAATAAATAATTAATTTGTTGAGAAAGGAAACTTC 941
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BYTV2 1096 bp DNA VRL 30-OCT-1994
LOCUS Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes,
DEFINITION complete cds.
ACCESSION L32167
VERSION L32167.1 GI:520791
KEYWORDS stem loop.
SOURCE Banana bunchy top virus DNA.
ORGANISM Banana bunchy top virus

```


Viruses; ssDNA viruses; Circoviridae.
 1 (bases 1 to 1096)
 Wu, R.-Y., You, L.-R. and Soong, T.-S.
 Nucleotide sequences of two circular single-stranded DNAs
 associated with banana bunchy top virus
 unpublished (1994)
 JOURNAL
 FEATURES
 source
 Location/Qualifiers
 1..1096
 /organism="Banana bunchy top virus"
 /db_xref="taxon:12585"
 1..7
 TATA_signal
 stem_loop
 gene
 71..928
 /gene="V2"
 71..928
 /gene="V2"
 /codon_start=1
 /protein_id="AA51426.1"
 /db_xref="GI:520792"
 /translation="MSSFKWCFTLNYSSAAREDFLALLKEEDVHYVVGDEVAPATG
 QKHLQYLKLSIRLGLKKYKGSRAHWEIAKGSDEONRRYCSKETILVLELGPVVP
 GSKRKLDRPESPEELKMDPSKYRCLAVESIKDARKINSEWHLEKEMONKLIQH
 IEGVDPDRSLINVGPEGEGKSTFARLSLKGWGIYNGKTSMDMHIITMDPDNHW
 IIDIPRSHDYLNVGVIDQIKNRVLINTKYEPVIRKDGQNVHVIVMANVLDPYKILS
 ERIKILNC
 85..309
 /gene="V1"
 85..309
 /gene="V1"
 /codon_start=1
 /protein_id="AA51427.1"
 /db_xref="GI:520793"
 /translation="MVLSELFRLSGARRLSRSEGGCSLLCRRRBSRSHRPEAPP
 GISPEKINSRRIKEVWLPLSLGDCERK"
 complement(544..699)
 /gene="C2"
 complement(544..699)
 /gene="C2"
 /codon_start=1
 /protein_id="AA51428.1"
 /db_xref="GI:520794"
 /translation="MMLCISDVFPPLIYPHPGPNDRYLAKVDFPPLGPTQMILRS
 SGTFSMC"
 complement(883..1059)
 /gene="C1"
 complement(883..1059)
 /gene="C1"
 /codon_start=1
 /protein_id="AA51429.1"
 /db_xref="GI:520795"
 /translation="MGLRPKNRFINYNDFLHMLIACSASRYRGSTCFILACGGSFLS
 QQLTIFILSSEILQ"
 347 a 231 c 244 g 274 t

Query Match
 Best Local Similarity 20.8%; Score 116.4; DB 58; Length 1096;
 Matches 305; Conservative 0; Mismatches 261; Indels 12; Gaps 2;
 QY 6 tcttgagtcacaaagccacgattggcccaacgattgtgtgaggaacctgatgaactccg 65
 DB 373 TCCTGGTTTCGAAGAGCGCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 432
 QY 66 ccttgagacccagcggatccacgaatgcgtgtgtacacgagcttcggtggaatggac 125
 DB 433 GATGACCATCCATCCCAAGATTCACAGATGCTTGGCAGTGGCAATCAATTAAGATGCCAG 492
 QY 126 aagatggccgcgtgaaatccgttccattccatatacaaatggcagcttgaagtgc 185
 DB 493 AATTAATTCGAATGGGTTTCAGCACTA-----AAAGATGCCAAATAAATAAT 543
 QY 186 gtctgcgacgagagccagcgcgacgatccacaaatctctgtgatatcgcgacgagacgg 245

Db 544 TCACACATCGAAGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGGTCCCAACGG 603
 QY 246 agggacgggaagtcocgtgtttgccaataatctcggaactcaagcccgactgggtctacac 305
 Db 604 AGGCGAAGAAAGTCAACCTTCGCAAGATATCTATCATTAATAAACCCGGATGGGATATAT 663
 QY 306 atggttggaaccagaaagacgtattgtaccagtagcatcgagaccacaaacgaaattt 365
 Db 664 CAACGGTGGAAAGACGTCGGATATGATGCATCATATAACGATGGATCCTGATAATCATTG 723
 QY 366 aatcctcgatgataccagggtgaatttagagattttaaattatgcctgttagaattgt 425
 Db 724 GATTATTGATATCCCAAGAGTCAATCAGATTATCTGAATTATGGCTTATAGAACAAAT 783
 QY 426 taagacagggcattccagttcggacaaaatacgaacc---ccttagttatcttgggttga 482
 Db 784 TAAGATAGAGTTTTATAATAACAAATACGAACCATGTGTGATTAGAAAAAGATGGACA 843
 QY 483 ccattgcatgtactogtatttcccgaatgctcctgcctgattatttgaataatcagcagga 542
 Db 844 AATGTCCATGTAATGTTATGTCGAATGTGTGCTGCTGATTATTGTAAATTTTCAGAGA 903
 QY 543 cagaataaaactcggaataatttaaagtagtattgttcac 580
 Db 904 TAGAATAAAATAAATAATTAATTTGTGAGAAAGAACTTC 941

RESULT 6
 AR063453
 LOCUS AR063453 1091 bp DNA PAT 29-SEP-1999
 DEFINITION Sequence 5 from patent US 5846705.
 ACCESSION AR063453
 VERSION AR063453.1 GI:5992761
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 1091)
 Wu, R., You, L. and Soong, T.
 Nucleotide sequence of two circular ssDNA associated with banana
 bunchy top virus and method for detection of banana bunchy top
 virus
 JOURNAL Patent: US 5846705-A 5 08-DEC-1998;
 FEATURES Location/Qualifiers
 SOURCE I..1091
 BASE COUNT 343 a 224 c 248 g 276 t
 ORIGIN

Query Match
 Best Local Similarity 19.7%; Score 114.8; DB 9; Length 1091;
 Matches 304; Conservative 0; Mismatches 262; Indels 12; Gaps 2;
 QY 6 tcttgagtcacaaagccacgattggcccaacgattgtgtgaggaacctgatgaactccg 65
 Db 368 TCCTGGTTTCGAAGAGCGCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 427
 QY 66 ccttgagacccagcggatccacgaatgcgtgtgtacacgagcttcggtggaatggac 125
 Db 428 GATGACCATCCATCCCAAGATTCACAGATGCTTGGCAGTGGCAATCAATTAAGATGCCAG 487
 QY 126 aagatggccgcgtgaaatccgttccattccatatacaaatggcagcttgaagtgc 185
 Db 488 AATTAATTCGAATGGGTTTCAGCACTA-----AAAGATGCCAAATAAATAAT 538
 QY 186 gtctgcgacgagagccagcgcgacgatccacaaatctctgtgatatcgcgacgagacgg 245
 Db 539 TCACACATCGAAGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGGTCTCGCCGG 598
 QY 246 agggacgggaagtcocgtgtttgccaataatctcggaactcaagcccgactgggtctacac 305

MEDLINE	99094637
REFERENCE	2 (bases 1 to 1015)
AUTHORS	Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetten,H.J. and Gronenborn,B.
TITLE	A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
JOURNAL	J. Virol. 73 (12), 10173-10182 (1999)
MEDLINE	20027244
REFERENCE	3 (bases 1 to 1015)
AUTHORS	Katul,L
TITLE	Direct Submission
JOURNAL	Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. biol. Sicherheit, Biologische Bundesanstalt f. Land-u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
FEATURES	Location/Qualifiers
source	1..1015 /organism="faba bean necrotic yellows virus" /virion /isolate="Egyptian EVI-93" /db_xref="taxon:59817" /lab_host="Vicia faba"
stem_loop	1..33
gene	60..1014 /gene="C7-Eg"
CDS	60..911 /gene="C7-Eg" /function="replication-associated protein" /notes="33.05 kda" /codon_start=1 /evidence=experimental /product="Rep protein" /protein_id="CA84025.1" /db_xref="GI:4995168" /db_xref="SPREMBL:O9UK0" /translacion="MPSIRATHWCTLFNPFSSIPENITADVOYISQHERVGHDLHGYIOXNNHVTLKKMKLLPGLAHLEMAKAPKAIEYCQKKSAIAGPWEIGTWISSGHSKRKLKRVQMKKIRESCTWNFDLRPWDELLKTVEEDPTLIWYGPHGGEGSAFAKYLTLKGEWMVTAGKATDMLYSLDPTCHVICIDTPRCRTREYNYSVIEQIKNRVIINTKYEPCITRDDGHNVHIVFCNFPLDPVTRISDDRIKLING"
TATA_signal	979..984 /gene="C7-Eg"
polyA_signal	1009..1014 /gene="C7-Eg"
BASE COUNT	346 a 168 c 228 g 273 t
ORIGIN	
Query Match	17.8%; Score 104; DB 58; Length 1015;
Best Local Similarity	56.9%; Pred. No. 8.4e-20;
Matches	Conservative 0; Mismatches 170; Indels 6; Gaps 2;
Qy	169 tggcagctgaagtgtcttgcacatcgaggacgcaggcagcatcgcaaatccctctgg 228
Db	507 TGCAGCATGAATTGTGTGAAGCAGTGTGAACAGAACCATGCCGTACAATTTTGG 566
Qy	229 atatcggcagcagcagcagcagcagcagggaagtcggttttgccaataatctcgactcaag 288
Db	567 GTGTATGCCCCACATGGAGAGAAGTAAGCGCGTTTGCGAAATACTTTAAOCCTTGAAG 626
Qy	289 ccogactggttctcacatgtgtgtggaaccgaaaggacgtattgtaccagtcacatcgag 348
Db	627 GAAGGTTGGTGTATACACTGGAGGAAAGCGACAGATATGCTTTTACAGCTATTCTATG 686
Qy	349 gaccacaaacgaatttaacctcgatgtaccaccagtgct---aatttagagttataaat 405
Db	687 GACCCCACATGCCACGTATTCATTTATCCACAGTGTACACAGAGAAGATAATATAAT 746
Qy	406 tatgccctgttagaatgtttaagaacaggggatcagttcggacaataacgacccctt 465
Db	747 TATTCTGTAATTGAACAGATATAAATAGAGTTATTATTAACAATAATGAACCTTGT 806

QY	466	agttatcttgggttgcg---accatgtgcagctactcgtatttgccaatgctcctgcctgat	522
Db	807	ACTATTAGAGACGATGGTCATAATGCTTCATGTAATGTTATTTGTAATTTTTTACCTGAT	866
QY	523	tatttgaataatcagcaggagcagaataaaaactgtggaattatttaagt	570
Db	867	GTAACAAGAAATCAGTGAAGATAGATAAAATAATAATAATGTTGAAAT	914

RESULT	11
AF216221	1109 bp DNA VRL 08-MAR-2001
LOCUS	Banana bunchy top virus satellite S1 replication initiation protein (ORF V1) gene, complete cds.
DEFINITION	
ACCESSION	AF216221
VERSION	AF216221.1 GI:12004325
KEYWORDS	Banana bunchy top virus.
SOURCE	Banana bunchy top virus
ORGANISM	Viruses; ssDNA viruses; Circoviridae.
REFERENCE	1 (bases 1 to 1109)
AUTHORS	Horser,C., Harding,R. and Dale,J.
TITLE	Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein
JOURNAL	J. Gen. Virol. 82 (Pt 2), 459-464 (2001)
MEDLINE	21102983
PUBMED	11161286
REFERENCE	2 (bases 1 to 1109)
AUTHORS	Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.
TITLE	Additional rep-encoding DNAs associated with banana bunchy top virus
JOURNAL	Arch. Virol. 146 (1), 71-86 (2001)
REFERENCE	3 (bases 1 to 1109)
AUTHORS	Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.
TITLE	Direct Submission
JOURNAL	Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia
FEATURES	Location/Qualifiers
source	1..1109
	/organism="Banana bunchy top virus"
	/isolate="Taiwan"
	/db_xref="taxon:12585"
	/note="satellite S1"
stem_loop	3..33
	/note="putative sequence"
mRNA	63..917
	/gene="ORF V1"
gene	63..917
	/gene="ORF V1"
CDS	63..917
	/gene="ORF V1"
	/codon_start=1
	/product="replication initiation protein"
	/protein_id="AAG44003.1"
	/db_xref="GI:12004326"
	/translations="MSSFKWCFPLNYSSAERDFLALKKEELNVAVVGDEVAVSSG OKHGLVSLKSGIKLGLKKYKSSRAHWERAGSDNARYCSKETLLELLELGFPAAS GNSRKLSEWSPERMTEQPEIYHRTYSVKLKKFEEFVHPCLDPRWQIQLETA IDEPDDRSIIWYGPNGNECKSTYAKSLMKDWFTYRGKKENILFSYVDGSEKHI VDFIKQCNILNDIVTEALKDRVIESTKYKPKLVELINIHVIVNANMPMEFKCISE DRKIIVYC"
BASE COUNT	346 a 213 c 257 g 293 t
ORIGIN	

Query Match	17.8%	Score 103.8;	DB 58;	Length 1109;
Best Local Similarity	57.1%	Pred. No. 9.7e-20;		
Matches 230;	Conservative	0;	Mismatches 167;	Indels 6;
Gaps				2;

169 tggcagcttgagtgctgtctggatcggagagcagcgagcgtatcgacacaatcctctgg 228
 QY
 519 TGGCAGATTCGAATTGACGGAGGCAATTGACGAGGAAACCCGATGATCGAAGCATCATCTGG 578
 dh

us-09-462-955-1_copy_409_991.rge

Thu Oct 25 13:08:07 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:35 ; Search time 5701.85 Seconds
(without alignments)
2259.729 Million cell updates/sec

Title: US-09-462-955-1_COPY_383_1215

Perfect score: 833

Sequence: 1 ggggtccttcagcacgga.....accatccagtgagtagct 833

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: gb_bal:*
17: gb_ba2:*
18: gb_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vil2:*
60: gb_vil3:*
61: gb_vil4:*
62: gb_vil5:*
63: gb_vil6:*
64: gb_vil7:*
65: gb_vil8:*
66: gb_vil9:*
67: gb_vil10:*
68: gb_vil11:*
69: gb_vil12:*
70: gb_vil13:*
71: gb_vil14:*
72: gb_vil15:*
73: gb_vil16:*
74: gb_vil17:*
75: gb_vil18:*
76: gb_vil19:*
77: gb_vil20:*
78: gb_vil21:*
79: gb_vil22:*
80: gb_vil23:*
81: gb_vil24:*
82: gb_vil25:*
83: gb_vil26:*
84: gb_vil27:*
85: gb_vil28:*
86: gb_vil29:*
87: gb_vil30:*
88: gb_vil31:*
89: gb_vil32:*
90: gb_vil33:*
91: gb_vil34:*
92: gb_vil35:*
93: gb_vil36:*
94: gb_vil37:*
95: gb_vil38:*
96: gb_vil39:*
97: gb_vil40:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	833	100.0	1291	58	CFDCG
2	132	15.8	1004	59	M29963 Coconut fol
3	132	15.8	1007	58	AJ005966 faba bean
4	120.6	14.5	1096	9	AJ132187 faba bean
5	120.6	14.5	1096	58	AR063452 Sequence
6	117	14.0	1091	9	BYTV2
7	115.6	13.9	1095	58	L32167 Banana bunc
8	108.6	13.0	1111	58	AR063453 Sequence
					AF216222 Banana bu
					U12586 Banana bunc

TRPGVKRPLAQRFAEPEDELRLEDGGYRCVHVHGAASVETWRMAAENPPFPYHNQ
LEVLISAIGPADDRITLWTCGDKGKSVFAYLGLKPDWFTYCGGTRKDVLYOIE
DKRNLILDPNLELYNALLCEVKNRATSSDKYEPYSLGFDHVLVFNVLDP
YKISHDRILKLNLI
314. .775
/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSPVPTNRIETTVRRNGCFSTESRLVLESGHDWPN
DLRLNMSAWKQADTEDALYELRWNGDGLKIRSHFHTITIGSLKCCILRSQSRT
IAOSSGYADETEETSPCLPILSDSPGSHVPEPTYCTSTRTQNEI"
complement(422..568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFQRPILSIPPLKRVQIRLGLPGVHVPPQIVGP
IVAF"
639. .797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MTRRRRRREVRCQISRTQARLVLMHWNQKGRIVPVHRGPKTK
FNPRCTQV"
complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MHTLNIPOFYSVPADFOIIRQIDIGKYEYMHVMEPKITKGFVVF
RTECPVLNTF"
1098. .1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVNGGPTIKDSIWIRTNLLILCLOCTQPLSTPIQVSSLLEKK
AASLYLPSICFCAIGRLS"
BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

U12587 Banana bunc
AB009047 Milk vetc
AF216221 Banana bu
AJ132185 faba bean
AR063451 Sequence
L32166 Banana bunc
AJ005964 faba bean
AB000922 Milk vetc
U16735 Subterranea
AB000921 Milk vetc
U02312 Banana bunc
X80879 FBMYV gene
AJ005968 faba bean
U16731 Subterranea
AB000920 Milk vetc
AJ132345 Nanovirus
AJ238493 Nanovirus
AJ132344 Nanovirus
AR102030 Sequence
AR102027 Sequence
U18077 Banana bunc
AR102025 Sequence
AR102028 Sequence
AR102040 Sequence
AF102780 Banana bu
S56276 orf VI {com
AR102032 Sequence
AR102034 Sequence
AR102029 Sequence
AR102031 Sequence
AF238874 Banana bu
U97525 Banana bunc
AR102033 Sequence
AR102026 Sequence
AF238875 Banana bu
AF246123 Banana bu

Query Match 100.0%; Score 833; DB 58; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.5e-227; Mismatches 0; Gaps 0;
Matches 833; Conservative

QY 1 ggtgtctctcagcagcagctccgactcgtcctggagtcacaaaggccacgattggccc 60
Db 383 GGGTGTCTTCGAGCAGGAGTCCCGACTCGTCTGGAGTCAAAAGGCCACGATTGGCCC 442

QY 61 aacgattgtcaggaaacctgatgaactccgcctggaagaccacccgaggtaccagagat 120
Db 443 AACGATTGTCTGAGGAACCTGATGAACCTCCGCTGGAAGACCCAGCGGATACCGAAGAT 502

QY 121 gcgtgtacacgagcttcggtggaatgacaaagatggccgctgaaatccgttcccat 180
Db 503 CGGTTGTACACGAGGCTCGGTGGAATGGACAGATGGGCCCTGAAATCCGTTCCCAT 562

QY 181 ttccatcacaaatggcagcttggaagtgtcgtcgcgacgagagcagcagcagcagc 240
Db 563 TTCCATATACATATGCGACCTTGAAGTGTCTGCGATCGGAGACCGAGCGAGCATC 622

QY 241 gcacatctcttgatgacgagcagcagcagcagcagcagcagcagcagcagcagc 300
Db 623 GCACATCTCTTGATATGCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 682

QY 301 atctcgaactcaagccgactggtctacacatgtgttggaacacgaagagcagcagc 360
Db 683 ATCTCGGACTCAAGCCGACTGTTCTACACATGTGTGTGAACCCAGAAAGAGCAGTATTGT 742

QY 361 accagtcacatgcaggagccacacacacacacacacacacacacacacacacacac 420

ALIGNMENTS

RESULT 1

LOCUS 1291 bp ss-DNA circular VRL
DEFINITION Coconut foliar decay virus, complete genome.
ACCESSION M29963
VERSION M29963.1 GI:323306
KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
ORGANISM
Viruses; ssDNA viruses; Nanovirus.

REFERENCE
1 (bases 1 to 1291)
Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
Nucleotide sequence of a circular single-stranded DNA associated
with coconut foliar decay virus
Virology 176, 648-651 (1990)

JOURNAL
MEDLINE
COMMENT Draft entry and printed sequence for [1] kindly submitted by
W.Rhode, for release after publication.
Location/Qualifiers
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40..70
/note="stem-loop structure"
103..975
/note="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSFRWCFTLNVEFEAAANVVRRIEISLNLVIAVIGDEVAPS
TQRLHGLGFIHLKTLGRRLQGLKTVLGNDRILHLEPTKSGEQNRDYSKERVLLHGVF

stem_loop
CDS

```

Db 743 ACCAGTACATCAGGACCCAAACAAATTTAATCCTCGATGTACCCAGGTGTAATTAG 802
QY 421 agtatttaattaccctgttagaagtgttaagaacagggaattcagttcggacaat 480
Db 803 AGTATTAAATATGCCCTGTAGAAATGTGTAAGAACAGGGCATTTCAGTTCGACAAAT 862
QY 481 acgaaccccttagttatcttgggttcgaccatgtgcagtactcgtatttgcacaatgtcc 540
Db 863 ACCAACCCCTTAGTATCTCGGTTCGACCAATGTGCAATGACTCGTATTTGCCAAATGTCC 922
QY 541 tgcctgattattgaataatcagcaggagacagaataaaactgtggaattatttaagtatgt 600
Db 923 TGCCTGTATTTTGAATATCAGCAGGACACAGATAAAACTGTGGAATATTAAAGTATGT 982
QY 601 gtcatctaaattacaccaataaccgcccgcacccagcgcgtactcgtttacattctatgaata 660
Db 983 GTCATCTAAATACACCAATACCCGCCGCCGCGGCTATCGCTTTACATCTTTATGAATA 1042
QY 661 tctgcccaggccgaagcctggaggtgctaccgcccgaagccggaacaataatgaa 720
Db 1043 TCCTGCCAGGCCGAAGCCTGGAGGTGCTACCCGCCGAGGCCGGAACAATATGAA 1102
QY 721 tcgagttatggcggcgcacaaataaaagattccatttggataagaacgaattctgtact 780
Db 1103 TCGAGTTATGGCGGGGCCCAATAAAGATTCCATTTGGTAAGAACCAATCTGTACT 1162
QY 781 ttgcttcagtcgacgaacacactttccacgtccaccaatccaggtagtagct 833
Db 1163 TTGCTTCAGTCGACGCAACCACTTTCCAGCTCACCACATCCAGGTGAGTAGCT 1215

```

```

RESULT 2
NV5966
LOCUS
DEFINITION
  faba bean necrotic yellows virus C9 gene.
  13-JAN-1999
ACCESSION
  AJ005966
VERSION
  AJ005966.1 GI:3550532
KEYWORDS
  C9 gene; component 9; putative; rep protein; replication associated
  protein.
SOURCE
  faba bean necrotic yellows virus.
  faba bean necrotic yellows virus.
  Viruses; ssDNA viruses; Nanovirus.
  1 (bases 1 to 1004)
  Katul,L.
  Direct Submission
  Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und
  Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und
  Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY
  2 (bases 1 to 1004)
  Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
  Ten distinct circular ssDNA components, four of which encode
  putative replication-associated proteins, are associated with the
  faba bean necrotic yellows virus genome
  J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
  99094637
FEATURES
    Location/Qualifiers
    1..1004
    /organism="faba bean necrotic yellows virus"
    /viral
    /isolate="SV292-88"
    /db_xref="taxon:59817"
    73..918
    /gene="C9"
    73..918
    /gene="C9"
    /function="putative replication associated (rep) protein"
    /codon_start=1
    /product="component 9"
    /protein_id="CA06789.1"
    /db_xref="GI:3550533"
    /db_xref="SPTREMBL:O91252"
    /translation="MSAVNWVETLNFAGEVPVLSFDERVOYAVWQHVRVNDHDIQGV

```

```

OLKKAKMNTVKNLIGNPHLEKMGSI EASAYAOKEESRVAGPWSGELLKKGSHK
RKIMELIKDPENLESEPOKVRAMANSAMDESRLAEEGFFMYFSWQETVIGLEE
EPNDRTIIWYGPNGNEGKSOQKFLGLKKDYILDGGKTQDMTYMLKKRPNRYVMD
IPRCNSELNYQFMELIKNRTIYSYKEPVGCIINNKIHVILANVLDPYKISQDRI
KIYC"
BASE COUNT      352 a      159 c      226 g      267 t
ORIGIN
Query Match
Best Local Similarity      15.88;      Score 132;      DB 59;      Length 1004;
Matches 270;      Conservative 0;      Mismatches 230;      Indels 0;      Gaps 0;
QY 94 tggaaagaccagggcgagataccgaagatgcgtgttacacggagcctcgggtggaatggacaa 153
Db 419 TGGAAAGAACCCCAAGAAATACAGAGAGCTATGCTTGGTCCGCCATGGGACGAATCTCGGA 478
QY 154 gatgggcccgtgaaatcccgcttcccatttccatatacacaattggcagcttgaagtgtgt 213
Db 479 AGCTTGGCGGAAGAGAAAGGCTTTCCCTATATGTTTACAGCTGGCAAGAAACAGTGTGG 538
QY 214 ctgcgacgcgagaccagcgagcgcgcacacacacacacacacacacacacacacacacac 273
Db 539 GCCTATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
QY 274 gagacgggaagtcggtgttggccaaatatactcgactcaagccgactggttctacacat 333
Db 599 ATGAAGCAAAATCACAGTTTGGTAAATTCCTGGGATTAAAAAAGATTACCTTTATTATAC 658
QY 334 gtggtgaaacagaaagagcgatgtgaccagtcacacacacacacacacacacacacacac 393
Db 659 CTGAGGTAAACCCCAAGATATGACATATGTTTAAAGAAAAATCCAAAGGCAAAATGTTG 718
QY 394 tctctgagtcacaggtgttaatttagagattttaaattatgcccgtgtagaattgttta 453
Db 719 TGATGGATATTCCTCGTCTGTAATTCGTAAATTCCTGGGATTAAAAAAGATTACCTTTATT 778
QY 454 agaacagggcattcagttcggcaaaatacgaaccccttagttatcttgggttcgaccatg 513
Db 779 AAAATAGAACCATATATAGTTATAATATGAACACCATGGATGTTATAATAATAATAAAA 838
QY 514 tgcattactcgtatttgcgaatgctcctgcctgattattgaaatacagcaggagacagaa 573
Db 839 TACATGTAATTTGATTAGCTAATGATTGCTTCCCTGATTGAAAAAATTAGTCAGGATAGAA 898
QY 574 taaactcgggaatatttaa 593
Db 899 TTAATAATAATTTATTGTTAA 918

```

```

RESULT 3
FBEL32187
LOCUS
DEFINITION
  faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian
  EVI-93.
ACCESSION
  AJ132187
VERSION
  AJ132187.1 GI:4995171
KEYWORDS
  C9-Eg gene; rep protein.
  faba bean necrotic yellows virus.
  faba bean necrotic yellows virus.
  Viruses; ssDNA viruses; Nanovirus.
  1 (bases 1 to 1007)
  Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
  Ten distinct circular ssDNA components, four of which encode
  putative replication-associated proteins, are associated with the
  faba bean necrotic yellows virus genome
  J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
  99094637
FEATURES
    Location/Qualifiers
    1..1007
    /organism="faba bean necrotic yellows virus"
    /viral
    /isolate="SV292-88"
    /db_xref="taxon:59817"
    73..918
    /gene="C9"
    73..918
    /gene="C9"
    /function="putative replication associated (rep) protein"
    /codon_start=1
    /product="component 9"
    /protein_id="CA06789.1"
    /db_xref="GI:3550533"
    /db_xref="SPTREMBL:O91252"
    /translation="MSAVNWVETLNFAGEVPVLSFDERVOYAVWQHVRVNDHDIQGV

```

JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 J. Virol. 73 (12), 10173-10182 (1999)
 20027244
 3 (bases 1 to 1007)
 Submitted (09-JAN-1999), Katul L., Inst. f. pflanzenvirologie,
 Mikrobiol., u. biol. Sicherheit, Biologische Bundesanstalt f. Land-
 u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
 Location/Qualifiers
 1. .1007
 /organism="faba bean necrotic yellows virus"
 /viroion
 /isolate="Egyptian EV1-93"
 /db_xref="taxon:59817"
 /lab_host="Vicia faba"
 1. .39
 /gene="C9-Eg"
 73. .1003
 /gene="C9-Eg"
 73. .918
 /gene="C9-Eg"
 /function="replication-associated protein"
 /note="32.53 kDa"
 /codon_start=1
 /evidence=experimental
 /product="Rep protein"
 /protein_id="CAB44027.1"
 /db_xref="GI:4995172"
 /db_xref="SPTREMBL:O9WIK2"
 /translation="MSAVNVEVTLNFAEGPEVLSFDERVOYAVQWHERVNHDIQCVI
 QLKRRKNTVNIIGNPHLEKMKGSIEBASAYAKOEKSRVAGPMSYCELLKKGSHK
 RKIMELIKDPNELEPEPKYRRAMAWAMDESKLAEBSGFPVTLYSQWETVLGLLE
 EPNDRIITWYTPNGNEKSGQFGLGLKKDYLPGGKTQDMTILMLKPNKRVNMD
 IPRCNSELYNQFMELIKNRTIFSYPYKPEVPGVCIINNKIHVILVLANLPDYELISQDRI
 KIIYC"
 939. .944
 /gene="C9-Eg"
 998. .1003
 /gene="C9-Eg"
 /gene="C9-Eg"
 353 a 161 c 227 g 266 t
 BASE COUNT
 ORIGIN
 TATA_signal
 polyA_signal
 Query Match 15.8%; Score 132; DB 58; Length 1007;
 Best Local Similarity 54.0%; Pred. No. 7.9e-27;
 Matches 270; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
 Qy 94 tggagaccgagcgataccgaagatcggttgatcacacgagcttcggtggaatggacaa 153
 Db 419 TGGGAAGAACCCAGAAATACAGAAAGAGCGATGGCTTCCGCCATGGACGAATCTCGA 478
 Qy 154 gatggccgcgtgaaatccggttccatcttccatatacacaattggcagcttgaagtgctg 213
 Db 479 AGCTTGTAAGAGAGGAGGCTTCCCTTATACAGCTTTACAGCTGGCAAGAACAGTGTGG 538
 Qy 214 ctgcgacgagagccgagcgagcagatcgacaaatcccttgatcgacgagacgagcgag 273
 Db 539 GCCTATTAGAGAGAGACCCCAATGACCGTATTATTATTTGGGTCTACGCCCAATGTTA 598
 Qy 274 gagacgggaagtcggtgttgcacaaatctcggactcgaagccgactggttctacacat 333
 Db 599 ACGAAGGAAATACATGATTTGGTAAATTCCTGGGATTAATAAAGATTAACCTTTATTAT 658
 Qy 334 gtgtgaaacagaaagacgattgttaccagtcacatcgagggaccacaaacgaatttaa 393
 Db 659 CTGGAGGTAAACCCAGATATGACATATATGTTAATGAAATATCCAAAGGCAATGTTG 718
 Qy 394 tctcgatgtaccaggtgtaatttagagattattaaattatgcccgtttagaatgtgta 453
 Db 719 TGATGGATATCCCGTGTGTAATTCCTGAATATTTAAATTTATCAATTTATGGAATTA 778
 Qy 454 agaacaggggattcagtttcggaacaaatcacgaaccccttagttatcttgggttcgacatg 513
 Db 179 AAAATAGAACCATATTTAGTTATAAATATGAACAGTGGATGATTAATAATAATAA 838
 Qy 514 tgcattactctgtatttgccaatgctcctgattatttgaataatcaggggacagaa 573
 Db 839 TACATGTAATTTAGTATGCTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 898
 Qy 574 taaaactgtggaattattaa 593
 Db 899 TTAATAATAATTTATGTTAA 918
 RESULT 4
 AR063452 1096 bp DNA PAT 29-SEP-1999
 LOCUS
 DEFINITION Sequence 4 from patent US 5846705.
 AR063452
 ACCESSION
 VERSION AR063452.1 GI:5992760
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 1096)
 AUTHORS Wu, R., You, L. and Soong, T.
 TITLE Nucleotide sequence of two circular ssDNA associated with banana
 bunchy top virus and method for detection of banana bunchy top
 virus
 JOURNAL Patent: US 5846705-A 4 08-DEC-1998;
 FEATURES Location/Qualifiers
 source 1. .1096
 BASE COUNT 347 a 231 c 244 g 274 t
 ORIGIN
 Query Match 14.5%; Score 120.6; DB 9; Length 1096;
 Best Local Similarity 52.6%; Pred. No. 1.5e-23;
 Matches 317; Conservative 0; Mismatches 274; Indels 12; Gaps 2;
 Qy 7 ttctcgagcagcgagctcccgactcgtcctgagtgagtcacaaaggccacgattggcccaacgat 66
 Db 348 TTCTTGAACCTGGGTACTCGGTGGTTCCTGTCGAGAACGCAAGCTTCTCCTATAGAT 407
 Qy 67 ttgctgaggaacatgatgaactccgctggaagaccacgagcgagataccgaagatgcgtg 126
 Db 408 TCAGAGAGACCCCTGAGGAATTTGAAGATGGACGATCCATCCAAAGTATCGCAGATGCTGG 467
 Qy 127 tacacgagcttcggtggaatggacaaatggcgctgaaatccgctcccatlccat 186
 Db 468 CAGTGAATCAATTAAGATGCCAGAATTAATTCGGAATGGGTTCACGAACATA 520
 Qy 187 atcaaatggcagcttggaatgctgctgagtcgagcgagcgagcgagcgagcgagcgagcg 246
 Db 521 --AAAGAAATGGCAAAATAAATAATTAACACATCGAAGGTTCCTGATGATCGAAGTA 578
 Qy 247 tctctggatatacgagcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 306
 Db 579 TCATCTGGGTATACGCTCCCAACGGAGGCGAAGAAAGTCAACCTTCGCAAGATATCTAT 638
 Qy 307 gactcaagcccgactggttctcacacatggtgggaaccagaaagacgattgttaccagt 366
 Db 639 CATTAAACCCGGATGGGATATATCAACGGTGGAAAGACGTCGGATATATGTCACATCA 698
 Qy 367 acatcagggaccacaaacgaaatttaactctgagtcgagcgagcgagcgagcgagcgagcg 426
 Db 699 TAACGATGGATCTGATATATCATTTGGATTTATGATATCCCAAGATGATTCAGATTATC 758
 Qy 427 taaattatgccctgttagaatgttgaacacagggcattcagttcggacaaatcagac 486
 Db 759 TGAATTTATGCGTTATAGAAATAAATAAAGATAGATTTTAAATAATAATAATAATAATA 818
 Qy 487 c---ccttagttatcctggttcgacatgctgactgactgatttgcgaatgtcctgc 543
 Db 819 CATGTGTGATTAGAAAAGATGGCAAAATGTCATGTAATTTGTTATGGCAATGTGTTCC 878

```

QY 544 ctgattattgaaatcagcaggacagaataaaactgtggaataatttaaagtatgtgc 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 879 CTGATTATTGTAATAATTCAGAGATAGATAAAATAATAATTAATTGTTGAGAAAGGAAC 938
|||
QY 604 atc 606
||
Db 939 TTC 941

RESULT 5
BYTV2
LOCUS BYTV2 1096 bp DNA VRL 30-OCT-1994
DEFINITION Banana bunchy top virus (BBTV DNA II) V1, V2, C1 and C2 genes,
complete cds's.
ACCESSION L32167
VERSION L32167.1 GI:520791
KEYWORDS stem loop.
SOURCE Banana bunchy top virus DNA.
ORGANISM Banana bunchy top virus
REFERENCE 1 (bases 1 to 1096)
AUTHORS Wu, R.-Y., You, L.-R. and Soong, T.-S.
TITLE Nucleotide sequences of two circular single-stranded DNAs
associated with banana bunchy top virus
JOURNAL Unpublished (1994)
FEATURES
    Location/Qualifiers
        1..1096
            /organism="Banana bunchy top virus"
            /db_xref="taxon:12585"
            1..7
            /gene="V1"
            71..928
            /gene="V2"
            71..928
            /gene="V2"
            /codon_start=1
            /protein_id="AAA51426.1"
            /db_xref="GI:520792"
            /translation="MSFEKWCFTLNYSAAERDFLALLKEEDVHYSVVGDEVAPATG
            QKHQGLYSLSKSLRLGLKKYGRSHWEIAKGSDFONRRYCKSTLVLELGTVPVP
            GSKRKLDRFESPELKMDFSKYRCLAVESIKDARINSEWVHELKELWONKLLQH
            IEGVDDRSIIWYGPNGEGEKSTFYRLSLKPGWYINGKTSMMHLITMDPNHW
            IIDIPRSHSDYLNLYGVIEQIKRVLINTEKPCVIRKQGNVHIVMANVLPDYCKIS
            EDRIKINC"
            85..309
            /gene="V1"
            85..309
            /gene="V1"
            /codon_start=1
            /protein_id="AAA51427.1"
            /db_xref="GI:520793"
            /translation="MVLHSEFLRSGARLRSSEGGGCSLLCRRRRSGHRPEAPP
            GISLPEKINSPIRIEKEVWLPCLSGDCERK"
            complement(544..699)
            /gene="C2"
            complement(544..699)
            /gene="C2"
            /codon_start=1
            /protein_id="AAA51428.1"
            /db_xref="GI:520794"
            /translation="MMCIISDVFPPPLIYPHPCGNDRLAKVDFFSPPLGPYQMLRS
            SGTPTSMC"
            complement(883..1059)
            /gene="C1"
            complement(883..1059)
            /gene="C1"
            /codon_start=1
            /protein_id="AAA51429.1"
            /db_xref="GI:520795"
            /translation="MGLRKNKFINDFLHMLIACASVYRGSTCFLIACGGSFLS
            QQLIIFLSSLEIQ"
            347 a 231 c 244 g 274 t

BASE COUNT 347 a 231 c 244 g 274 t

```

```

ORIGIN
Query Match 14.5%; Score 120.6; DB 58; Length 1096;
Best Local Similarity 52.6%; Pred. No. 1.5e-23;
Matches 317; Conservative 0; Mismatches 274; Indels 12; Gaps 2;

QY 7 ttctcgagcagcggagtcctccgactcgtctctgtagtcaaaaggccacagatggcccaacgat 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 TTCATTGAACCTGGGTACTCCCGTGGTTCCTGTTCCGAAAGAGCCCAAGCTTCGATAGAT 407

QY 67 ttgctgaggaaacctgatgaactccgctggagacacccagcgagataccgaagatcgcttg 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 TCAGAGAGAGACCCCTGAGGAATTAAGATGACGATCCATCAAGTAGTATCGCAGATGCTTGG 467

QY 127 tacacggagcttcggtggaaatgacaagatggccgctgaaatccgttcccatctccat 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 CAGTGGGAATCAATTAAGATGCCAGAAATTAATTCGGAATGGGTTCACGAACATA----- 520

QY 187 atcaaatggcagcttgaagtgtctgtcgatcgagagccagcgagacgacgacgacacaa 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 521 --AAAGAATGGCAAAATAAATTAATTCACACATCGAAGGTGTTCTGTGATGATCGAAGTA 578

QY 247 tctctggatatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 306
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 TCACTGGGTATACGGTCCCAACGGAGGCGAAGGAAGTCAACCTTCGCAAGATATCTAT 638

QY 307 gactcaagcccgactgttctacacatgtgtgggaaccagaaagacgacatattgtaccagt 366
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 639 CATTAACACCCCGATGGGGATATATCAACGGTGGAAAGACGTCGGATATGATGCACATCA 698

QY 367 acatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 699 TAACGATGATCCTCATATAATCATTCGATATTCATATCCCAAGAGTCATTTCAGATTATC 758

QY 427 taattatgccctgttagaattgttaagaacagcagcagcagcagcagcagcagcagcagc 486
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 759 TGAATTATGGCGTTATAGACACAAATTAAGATAGAGTTTTTAATAACAAAATACGAAC 818

QY 487 c---ccttagttattcgtgggttcgacctgtgcattgtactgtatttgcgaatgctctgc 543
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 819 CATGTGTGATTAGAAAAGATGCACAAAATGTCATGTAATTTGTTATGCAAAATGTTTGC 878

QY 544 ctgattattgaaatcagcaggacagaataaaactgtggaataatttaaagtatgtgc 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 879 CTGATTATTGTAATAATTCAGAGATAGATAAAATAATAATTAATTGTTGAGAAAGGAAC 938

QY 604 atc 606
||
Db 939 TTC 941

RESULT 6
AR063453
LOCUS AR063453 1091 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 5 from patent US 5846705.
ACCESSION AR063453
VERSION AR063453.1 GI:5992761
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1091)
AUTHORS Wu, R., You, L., and Soong, T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana
bunchy top virus and method for detection of banana bunchy top
virus
Patent: US 5846705-A 5 08-DEC-1998;
Location/Qualifiers
    1..1091
        /organism="unknown"
        /organelle="mitochondrion"
JOURNAL
FEATURES
    source
        BASE COUNT 343 a 224 c 248 g 276 t
ORIGIN

```

1

1

RESULT 8

Query Match	13.0%;	Score 108.6;	DB 58;	Length 1111;
Best Local Similarity	57.0%;	Pred. No. 4e-20;		
Matches 239;	Conservative	0;	Mismatches 174;	Indels 6;
Gaps	2;			
QY	195	tgccagcttgaaagtcgtctgcgatcggagacgagccgagcgatcgcacaaatcctctgg	254	
Db	535	TGCAGATTCATTTGACGGAGGCAATTGACGAGGACCAGATGATCGAACGATCTCTGG	594	
QY	255	atctcgccagcgagcgagggagcggggaagtcgctgtttgccaaatactctcggaactcaag	314	
Db	595	GTCTATGTGTCGGAATGATGATGAGGGCAATCAACATATGCGAAGCTCAT	651	
QY	315	cccgaactgggtctcacacatgtgtggaaaccgaaaggacgtattgtaccagtatcat	371	
Db	652	AAGCACTGGTCTACACCGAGGTGGGAAGAGGAGAACATACTGTTCTCTTACGTGGAC	711	
QY	372	gaggaccaccaaaacgaaatttaactccctcgatgtaccacaggtgtaattlagagatttaaat	431	
Db	712	GAAGGATCTCAGAAAGCATATTGTTATTGATATCTCTGCTGTAAATCAGGATTAATTAAT	771	
QY	432	tatgccctgttagaalggtttaagaaacagggcattcagttcggacaaataacgaacccctt	491	

GYLOMKKOTLLKMKLELLPAGHLEMARAPKKAIDYCKOKKETAIDCPWEYGTWISTGSH
 KKKLMBEDDEEMKLEDPGLYRCLSRVMTKVRKNSWDYDLRWQDMLLTIEQ
 EDDRTILWYCPHGEGSKSPFAYLTLKEGWNVTAGKATDMLYSYSLDPTCHVICD
 IPRCKEYINAVIEIKNRILIIINIKTEPCTIRDHGHVIVFANYLPDVTRISED
 RIKIYYC
 887. .892
 /gene="Rep"
 898. .903
 /gene="Rep"
 BASE COUNT 348 a 179 c 232 g 263 t
 ORIGIN
 Query Match 12.7%; Score 105.4; DB 58; Length 1022;
 Best Local Similarity 51.5%; Pred. No. 3.3e-19;
 Matches 325; Conservative 0; Mismatches 291; Indels 15; Gaps 3;
 QY 12 gagcagcaggtcccgactcgtccctggagtgcaaaaggccagcagattgcccacacatttgc 71
 Db 330 GAGTATGGGACCTGGATTCTTACAGGTAGTCACAAACGAAAGCTTATGGAAGATTGAT 389
 QY 72 gaggaacctgagtaactccgctggaagacccagcggtatccaggaagatcgctgtgtacac 131
 Db 390 GAAGACCCAGAGAGATGCAAGCTTGAAGATCCAGGATTATACAGCGCATCTTT----- 442
 QY 132 ggagcttcgttggaatggacaagatggggccgctgaaataccggtccattcccatatcac 191
 Db 443 --ATCGGAGTTTCAGATGACTTAAAGTTAGAGAGAAGAAATCTTGGGATTACGATCTTCGT 500
 QY 192 aattggcagctgaagtctgtctgcagtcgagcagcgagcagcgagcagcagcagcagcagc 251
 Db 501 CCATGGCAGGATGAATTTGTTGAAGACGATTAACAGGACAGATGATCGTACTTATCTG 560
 QY 252 tggaatgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 311
 Db 561 TGGGTGATGGCCACATGAGGCGGAGGCGAAAGTGTCTTTGCTAAATACTTAACTTAACTTG 620
 QY 312 aagccgactggtttcacacatggttggaacccagcagcagcagcagcagcagcagcagcagc 371
 Db 621 AAGGAAGTTGGTGTATACAGCTGGAGGGAAGAGCAGCAGATATGCTTTACAGCTATCA 680
 QY 372 gagc 428
 Db 681 TTGACCCACATGCCACCTGATGATATACCCAGATGATACCCAGATGATACCCAGATGATAT 740
 QY 429 aattgacctgtagatggttgaagacagcagcagcagcagcagcagcagcagcagcagcagc 487
 Db 741 AATTACGCACTAATTGAGCAAAATTAATAATGAAATTAATAATGAAATTAATAATGAAATTA 800
 QY 488 --cctagttatcttggttgcacatgctgctgctgctgctgctgctgctgctgctgctgctg 545
 Db 801 TGCATATTAGACAGATGCTCATATGTCCTGTAATGTTGCTGTAATGTTGCTGTAATGTTGCT 860
 QY 546 gattatttgaatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 605
 Db 861 GATGTAACAAGGATTAGTGAAGACAGATAAATAAATAAATAAATAAATAAATAAATAAATA 920
 QY 606 ctaataac 636
 Db 921 TTAATAAACTACGCGAAGCGTATTGGACACGC 951
 RESULT 11
 AF216221
 LOCUS AF216221 1109 bp DNA VRL 08-MAR-2001
 DEFINITION Banana bunchy top virus satellite S1 replication initiation protein
 (ORF V1) gene, complete cds.
 ACCESSION AF216221
 VERSION AF216221.1 GI:12004325
 KEYWORDS Banana bunchy top virus.
 ORGANISM Banana bunchy top virus.
 SOURCE Viruses; ssDNA viruses; Circoviridae.

595 GTCATATGTCGCAATGTAATGAGGGCAAAATCAACATATCGAAGTCAIT---AATGAAG 651
 QY 315 cccgactggttctacacatggtggtggaacagcagcagcagcagcagcagcagcagcagcagc 371
 Db 652 AAGGACTGTTCTACACAGAGGTGGAAGAAAGGAGAACTACTGTCTCCTACGTGGAC 711
 QY 372 gaggaacccaaacgaatttaactcctcgatgaccagcagcagcagcagcagcagcagcagcagc 431
 Db 712 GAAGGATCTGAGACCATATTGTTGATTTGATTCCTCGCTGTAATCAGGATTTATTAAT 771
 QY 432 tatgcccctgttagaattggtttaagacagcagcagcagcagcagcagcagcagcagcagcagc 491
 Db 772 TATGATGTTATAGAGCATTAAGAGTAGGTGATAGAGAGTACTAAATATAAACCTATT 831
 QY 492 agttattcgtggttcgaccatgctgactgctgctgctgctgctgctgctgctgctgctgctg 551
 Db 832 AAGTTAGTTGAATGAATAATACATGCTAATTTGTCATGCTAATTTATTCATGCCGAATTC 891
 QY 552 ttgaataacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 611
 Db 892 TGTAAATCTCCGAAGAGATGAATAAGATTATTATTGTTAAATAACACGCTATGACAAT 951
 QY 612 tacac 616
 Db 952 CGTAC 956
 RESULT 10
 AB009047 1022 bp DNA circular VRL 30-OCT-1998
 LOCUS Milk vetch dwarf virus genome segment 10 encoding replication
 DEFINITION initiation protein, complete sequence.
 ACCESSION AB009047
 VERSION AB009047.1 GI:3798737
 KEYWORDS putative viral replication-initiation protein.
 SOURCE milk vetch dwarf virus (specific_host:Astragalus sinicus,
 lab_host:Pisum sativum) DNA.
 ORGANISM
 Viruses; ssDNA viruses; Nanovirus.
 REFERENCE 1 (bases 1 to 1022)
 AUTHORS Sano.Y.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-1997) to the DDBJ/EMBL/GenBank databases.
 Yoshitaka Sano, Kyoto Institute of Technology, Department of
 Applied Biology; Matsugasaki, Sakyo-Ku, Kyoto, 606, Japan
 (E-mail:ysano@ipc.kit.ac.jp, Tel:075-724-7764, Fax:075-724-7764)
 2 (sites)
 Sano.Y., Wada.M., Hashimoto.Y., Matsumoto.T. and Kojima.M.
 Sequences of ten circular ssDNA components associated with the milk
 vetch dwarf virus genome
 J Gen. Virol. 79 (Pt 12), 3111-3118 (1998)
 JOURNAL 99094638
 MEDLINE Location/Qualifiers
 FEATURES
 source
 1. .1022
 /organism="milk vetch dwarf virus"
 /specific_host="Astragalus sinicus"
 /db_xref="taxon:67585"
 /chromosome="segment 10"
 /lab_host="Pisum sativum"
 /note="isolated from sieve element of broadbean"
 1. .29
 stem_loop
 57. .908
 gene
 /gene="Rep"
 57. .908
 CDS
 /function="site-specific DNA nicking and joining"
 /note="putative viral replication-initiation protein"
 /codon_start=1
 /evidence=not_experimental
 /product="replication initiation protein"
 /protein_id="BA034048.1"
 /db_xref="GI:3798738"
 /translation="MPSIRAIHWCFTLNFSGKIPEIWTADVOYSIQHVRVNHDLQ

Qy	552	ttgaataatcagcgaggacagaataaaactgtggaatatatttaaagtatgtcatctaaaat	611
Dd	876	TGTAATAATCTCCGAAGATAGAATAAAGATTATTATTTCGTAAAAAGAAATTTATATTATG	935
Qy	612	tacaccaataaccgcgccgc	630
Dd	936	CACATGCACAATCGTACGC	954
 RESULT 12			
FBI132185			
LOCUS			
DEFINITION	FBE132185 1015 bp DNA circular VRL 16-SEP-2000		
	faba bean necrotic yellows virus C7-Eg gene, isolate Egyptian EVI-93.		
ACCESSION	AJ132185		
VERSION	AJ132185.1 Gi:4995167		
KEYWORDS	C7-Eg gene; rep protein.		
SOURCE	faba bean necrotic yellows virus.		
ORGANISM	faba bean necrotic yellows virus.		
	Viruses; ssDNA viruses; Nanovirus.		
REFERENCE	1 (bases 1 to 1015)		
AUTHORS	Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.		
TITLE	Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome		
JOURNAL	J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)		
MEDLINE	99094637		
REFERENCE	2 (bases 1 to 1015)		
AUTHORS	Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetten,H.J. and Gronenborn.B.		
TITLE	A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants		
JOURNAL	J. Virol. 73 (12), 10173-10182 (1999)		
MEDLINE	20027244		
REFERENCE	3 (bases 1 to 1015)		
AUTHORS	Katul,L.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-Jan-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. Land-u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY		
FEATURES	Location/Qualifiers 1..1015 /organism="faba bean necrotic yellows virus" /viral /isolate="Egyptian EVI-93" /db_xref="taxon:59817" /lab_host="Vicia faba" 1..33 60..1014 /gene="C7-Eg" 60..911 /gene="C7-Eg" /function="replication-associated protein" /note="33.05 kDa" /codon_start=1 /evidence=experimental /product="Rep protein" /protein_id="CAH44025.1" /db_xref="GI:4995168" /db_xref="SPTRMBL:Q9WIK0" /translation="MPSIRATHWCFTLNFSKSGEINWTADVOYSTWOHERVCHDLHG GYIOGNKHVTLLKKMLPGAHLEMAKPAPKEICOKESALAGPWEYGVTSWSGH KRKLREYDGDPEEMKLEDPLCYRCLRSVMAKKIRESCTWNFLDRPWDELLKTGSH EPDDRTLIWYGPHGGEGSAFAYILTLKEGWMTTAGKATMDLYSYSLDPDTCHVICID IPRCITREYINYSVIEIQIKNRVIINTKYECTIRDGDHNHVIVFCNLPDVTRISED RIKINC"		
TATA_signal	979..984 /gene="C7-Eg"		
polyA_signal	1009..1014 /gene="C7-Eg"		
BASE COUNT	346 a 168 c 228 g 273 t		

100

us-09-462-955-1_copy_383_1215.rge

Thu Oct 25 13:08:01 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:34 ; Search time 5701.85 Seconds
(without alignments)
1144.785 Million cell updates/sec

Title: US-09-462-955-1_COPY_583_1004

Perfect score: 422

Sequence: 1 cttgaagtgtgtctgcgat.....catctaaattacaccaatac 422

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_ba4.*
5: gb_ba5.*
6: gb_ba6.*
7: gb_ba7.*
8: gb_ba8.*
9: gb_ba9.*
10: gb_ba10.*
11: gb_ba11.*
12: gb_ba12.*
13: gb_ba13.*
14: gb_ba14.*
15: gb_ba15.*
16: gb_ba16.*
17: gb_ba17.*
18: gb_ba18.*
19: gb_ba19.*
20: gb_ba20.*
21: gb_ba21.*
22: gb_ba22.*
23: gb_ba23.*
24: gb_ba24.*
25: gb_ba25.*
26: gb_ba26.*
27: gb_ba27.*
28: gb_ba28.*
29: gb_ba29.*
30: gb_ba30.*
31: gb_ba31.*
32: gb_ba32.*
33: gb_ba33.*
34: gb_ba34.*
35: gb_ba35.*
36: gb_ba36.*
37: gb_ba37.*
38: gb_ba38.*
39: gb_ba39.*
40: gb_ba40.*
41: gb_ba41.*
42: gb_ba42.*
43: gb_ba43.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vi.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vi1.*
59: gb_vi2.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_ro1.*
95: gb_ro2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	422	100.0	1291	58	CFDCG
2	117.6	27.9	1007	58	FBE132187
3	116	27.5	1004	59	NYV5966
4	114.2	27.1	1096	9	AR063452
5	114.2	27.1	1096	58	BYTV2
6	112.6	26.7	1091	9	AR063453
7	112.4	26.6	1095	58	AF216222
8	103.2	24.5	1111	58	BBU12586

M29963 Coconut fol
AJ132187 faba bean
AJ005966 faba bean
AR063452 Sequence
L32167 Banana bunc
AR063453 Sequence
AF216222 Banana bu
U12586 Banana bunc

9	102.8	24.4	1127	58	BBU12587	U12587	Banana bunc
10	99.2	23.5	1015	58	FBE132185	AF132185	faba bean
11	98.4	23.3	1109	58	AF216221	Banana bu	
12	96.8	22.9	1014	59	NFV5964	AJ005964	faba bean
13	96.8	22.9	1106	9	AR063451	AF063451	Banana bunc
14	96.8	22.9	1106	58	BYTV1	L32166	Subterranea
15	95.4	22.6	1017	59	SCU16735	U16735	Subterranea
16	95	22.5	1002	58	FBNYYREP	XB00879	FBNY gene
17	94.6	22.4	1000	58	AB000922	Milk vetc	
18	94.4	22.4	1009	58	AB000921	Milk vetc	
19	94	22.3	1022	58	AB005047	AB009047	Milk vetc
20	93.6	22.2	1095	59	U02312	U02312	Banana bunc
21	93	22.0	1004	59	NTV5968	AJ005968	faba bean
22	92.8	22.0	1007	58	AB000920	AB000920	Milk vetc
23	90.6	21.5	1022	59	SCU16731	U16731	Subterranea
24	89.4	21.2	1375	58	CLE132345	AJ132345	Nanovirus
25	87	20.6	1367	58	AYE238493	AJ238493	Nanovirus
26	85.4	20.2	1376	58	CLE132344	AJ132344	Nanovirus
27	62.8	14.9	1110	9	AR010230	AR010230	Sequence
28	59.6	14.1	982	9	AR112007	AR112007	Sequence
29	59.6	14.1	1110	9	AR010227	AR010227	Sequence
30	59.6	14.1	1110	58	BBU18077	U18077	Banana bunc
31	59.6	14.1	1111	9	AR010225	AR010225	Sequence
32	59.6	14.1	1111	9	AR010228	AR010228	Sequence
33	59.6	14.1	1111	9	AR010240	AR010240	Sequence
34	59.6	14.1	1111	58	AF102780	AF102780	Banana bu
35	59.6	14.1	1111	59	S56276	S56276	orf V1 [com
36	58.6	13.9	1103	9	AR010232	AR010232	Sequence
37	58.6	13.9	1105	9	AR010234	AR010234	Sequence
38	58	13.7	1109	9	AR010229	AR010229	Sequence
39	58	13.7	1111	9	AR010231	AR010231	Sequence
40	57	13.5	1103	58	AF238874	AF238874	Banana bu
41	57	13.5	1103	59	U97525	U97525	Banana bunc
42	57	13.5	1104	9	AR010233	AR010233	Sequence
43	56.4	13.4	1111	9	AR010226	AR010226	Sequence
44	53.8	12.7	1103	58	AF238875	AF238875	Banana bu
45	53.2	12.4	1103	58	AF246123	AF246123	Banana bu

RESULT	1
CDCG	CFDCG
LOCUS	1291 bp ss-DNA circular VRL
DEFINITION	Coconut foliar decay virus, complete genome.
ACCESSION	M29963
VERSION	M29963.1 GI:323306
KEYWORDS	circular; complete genome.
SOURCE	Coconut foliar decay virus DNA, clones NMORG, PORG, and PORGSAU3A.
ORGANISM	Coconut foliar decay virus
REFERENCE	Viruses; ssDNA viruses; Nanovirus.
AUTHORS	1 (bases 1 to 1291)
TITLE	Ronde.W., Randles,J.W., Langridge,P. and Hanold,D. Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
JOURNAL	Virology 176, 648-651 (1990)
MEDLINE	90266484
COMMENT	Draft entry and printed sequence for [1] kindly submitted by W.Rhode, 15-WAR-1989, for release after publication.
FEATURES	Location/Qualifiers
source	1..1291
	/organism="Coconut foliar decay virus"
	/db_xref="taxon:12474"
stem_loop	40..70
	/note="Stem-loop structure"
CDS	103..975
	/note="ORF 1"
	/codon_start=1
	/protein_id="AAA42894.1"
	/db_xref="GI:323307"
	/translation="MGSSIRRCWCTLNAYETEEEAANYVRRESNLNYAIIVGDEVAPS CHNHAEHREPSQSDNPYCSKFRVLI.FHGVP

```

|||||
Db 943 AGCAGGACAGAAATAAAACCTGTGAATATTTAAAGTATGTGTCACTAAATACACCAAT 1002
QY 421 ac 422
Db 1003 AC 1004

RESULT 2
FBE132187
LOCUS FBE132187 1007 bp DNA circular VRL 16-SEP-2000
DEFINITION faba bean necrotic yellows virus C9-Eg gene, isolate Egyptian
ACCESSION AJ132187
VERSION AJ132187.1 GI:4995171
KEYWORDS C9-Eg gene; rep protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus.
REFERENCE 1 (bases 1 to 1007)
AUTHORS Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
MEDLINE 99094637
REFERENCE 2 (bases 1 to 1007)
AUTHORS Timchenko,T., de Kouchkovsky,F., Katul,L., David,C., Vetten,H.J. and Gronenborn,B.
TITLE A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants
JOURNAL J. Virol. 73 (12), 10173-10182 (1999)
MEDLINE 20027244
REFERENCE 3 (bases 1 to 1007)
AUTHORS Katul,L.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-1999) Katul L., Inst. f. Pflanzenvirologie, Mikrobiol., u. Biol. Sicherheit, Biologische Bundesanstalt f. Land- u. Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
FEATURES
    source
        1..1007
            /organism="faba bean necrotic yellows virus"
            /viroion
            /isolate="Egyptian EV1-93"
            /db_xref="taxon:59817"
            /lab_host="Vicia faba"
            1..39
            /note="32.53 kba"
            73..1003
            /gene="C9-Eg"
            73..918
            /gene="C9-Eg"
            /function="replication-associated protein"
            /evidence="experimental"
            /product="Rep protein"
            /protein_id="CA544027.1"
            /db_xref="GI:4995172"
            /db_xref="SPTREMBL:O9WIK2"
            /translation="MSAVNVVFTLNFAGEVPLSFDERVOYAVWCHERVNHDHTQGV
            QLKKAQMTVKNIIGGNPHLEKMKKSTEEASVAOKEESRVAGPWSYGETLLKGGSHK
            RKIMELIKDPNELEPEOKYRRAMAWSAMDESRLAEGGEPYTLISWQETVGLGLEE
            EPNDRIITWYGPNGECKSOGKFLGLKLDYLYLPGKTQDMTYMLMKPNKANVMD
            IPRCSEVLYNQFMELIKNRTIFSKYEPVGCIIINNKIHVILVANLPLDYKTSQDRI
            KIIYC"
            939..944
            /gene="C9-Eg"
            998..1003
            /gene="C9-Eg"
            353 a 161 c 227 g 266 t

TATA_signal
polyA_signal
BASE COUNT 353 a 161 c 227 g 266 t
ORIGIN

```

```

Query Match 27.9%; Score 117.6; DB 58; Length 1007;
Best Local Similarity 56.4%; Pred. No. 2.8e-23;
Matches 219; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 6 agtgctgtcgtcgatcgagaccagcgagcgatcgacacatctctctgatatcggaag 65
|||||
Db 531 AGTGTGGGCTATTAGAGAACAGACCCCAATGACCGTATTATTATTTGGGTCTACGGGCC 590
|||||

QY 66 agacgagggagagcggaaagtcctgtttgccaaatactcgcgactcaagcccgactggt 125
|||||
Db 591 AAATGGTAACGAAGAGAAATACAGTTTGTGTAATTCCTGGGATTAATAAAGATTAACCT 650
|||||

QY 126 ctacacatgtgtggaaccagaaagacgtattgtaccagtcacatcgaggagcccaaacg 185
|||||
Db 651 TTATTACCTGGAGGTAACCAACCATGATGACATATATGTTAATGAAAAATCCAAAGGC 710
|||||

QY 186 aaatttaactctcgatgtaccaggtgttaatttagagatttaaaattatgcctgttaga 245
|||||
Db 711 AAATGGTGTGATGGATATTCCTCGTTGTAATTCCTGAATATTAAATTAATCAATTATGGA 770
|||||

QY 246 atgtgttaagacagggcattcagtcgcgacaaatacgaaccccttagtattcttggtt 305
|||||
Db 771 ATTAATTAATAATAGAACCATATTTAGTTATAAATATGAACGAGTTGGATGATTATAAA 830
|||||

QY 306 cgaccatgtcgtactcgtattgtccaatgctcgtcgtcgtattatttgaataatcacag 365
|||||
Db 831 TAATAAAATACATGTAATGTTAGTAAATGTAATGTTAGTAAATGTTAGTAAATGTTAGTAA 890
|||||

QY 366 gcacagaataaaactgtggaatatttaa 393
|||||
Db 891 GCACAGATTAATAATTAATTTATTTGTTAA 918
|||||

RESULT 3
LOCUS NTV5966
DEFINITION faba bean necrotic yellows virus C9 gene. 13-JAN-1999
ACCESSION AJ005966
VERSION AJ005966.1 GI:3550532
KEYWORDS C9 gene; component 9; putative; rep protein; replication associated protein.
SOURCE faba bean necrotic yellows virus.
ORGANISM faba bean necrotic yellows virus.
REFERENCE 1 (bases 1 to 1004)
AUTHORS Katul,L.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11-12, 38104 Braunschweig, GERMANY
REFERENCE 2 (bases 1 to 1004)
AUTHORS Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.
TITLE Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome
JOURNAL J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)
MEDLINE 99094637
FEATURES
    Location/Qualifiers
        1..1004
            /organism="faba bean necrotic yellows virus"
            /viroion
            /isolate="SV292-88"
            /db_xref="taxon:59817"
            73..918
            /gene="C9"
            73..918
            /gene="C9"
            /function="putative replication associated (rep) protein"
            /product="component 9"
            /protein_id="CAA06789.1"
            /db_xref="GI:3550533"

```

RESULT	4
AR063452	DNA
LOCUS	1096 bp PAT
DEFINITION	Sequence 4 from patent US 5846705.
ACCESSION	AR063452
VERSION	AR063452.1 GI:5992760
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 1096)
AUTHORS	Wu,R., You,L. and Soong,T.
TITLE	Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus
JOURNAL FEATURES	Patent: US 5846705-A 4 08-DEC-1998; Location/Qualifiers 1..1096 /organism="unknown"
BASE COUNT ORIGIN	347 a 231 c 244 g 274 t
Query Match	27.1%; Score 114.2; DB 9;
Best Local Similarity Matches	57.5%; Pred. No. 2.7e-22; 225; Conservative 0; Mismatches 163; Indels 3; Gaps
OY	19 atcgagaggccggaatgcacactctcgttatatgccggaccgacgaggagac 78

CDS	complement(544..699) /gene="C2" /codon_start=1 /protein_id="AA51428.1" /db_xref="GI:520794" /translation="MMCIISDVFPLYPHPGFDNDRLAKVDFPSPPLGYPYQMILRS SGTFSMC" complement(883..1059) /gene="C1" complement(883..1059) /gene="C1" /codon_start=1 /protein_id="AA51429.1" /db_xref="GI:520795" /translation="MGLRPNKRFINYNDFLMILMIACASRYVGSTCFLIACGGSLFS QQLIIFLSSILQ"																
gene																	
CDS																	
BASE COUNT	347 a	231 c	244 g	274 t													
ORIGIN																	
Query Match	27.1%; Score 114.2; DB 58; Length 1096;																
Best Local Similarity	57.5%; Pred. No. 2.7e-22;																
Matches	225;	Conservative	0;	Mismatches	163;	Indels	3;	Gaps	1;								
QY	19	atcggagagcgcgagcgcgcacatctctcggatatcggacgagacgagcggagac	78														
Db	551	ATCGAAGGTGTCCTGATGATCGAAGTATCATCTGGGTATACGGTCCCAACGGAGCGAA	610														
QY	79	gggaagtcgcgtgttgccaaatatctcggactcaagcccgactggtctcacatgtggt	138														
Db	611	GGAAAGTCAACCTTCGCAAGATATCTATCATATAAACCCGGATGGGATATATCAACGGT	670														
QY	139	ggaaccgaaaggcgtattgtaccagtacatcgagaccaccccaaaacgaaatttaactc	198														
Db	671	GGAAAGCGTCGATATCATGCCATCATACGATGATCCTGATAATCATTTGGATTATT	730														
QY	199	gatgtaccaggtgtaatttagagtatttaaattatgccctgttagaattgtttaagaac	258														
Db	731	GATATCCCGAGAGTCATTCAGATTATCTCAATTATGGCGTTATAGACAAATTAGAAT	790														
QY	259	agggcattcagtcgcgacaatacgaacc---cctagtatcttgggttcgaccatgtg	315														
Db	791	AGAGTTTATAATATCAATAAGAACCAACATGTCGATTAGAAAAATGGACAAATGTC	850														
QY	316	catgtactcgtatttgccaatgctcgtccctgattattgaaatcagcagggacagaata	375														
Db	851	CATGTAATTGTTATGGCAATGCTTGCCCTGATTATTGTAATAATTTCAGAAGTAGAATA	910														
QY	376	aaactgtggaattttaaagtatgtgtcatc	406														
Db	911	AAATAATTAAATTGTTGAGAAAGGAACTTC	941														
RESULT	6																
AR063453																	
LOCUS	AR063453	1091 bp	DNA	PAT													
DEFINITION	Sequence 5 from patent US 5846705.																
ACCESSION	AR063453																
VERSION	AR063453.1 GI:5992761																
KEYWORDS																	
SOURCE	Unknown.																
ORGANISM	Unknown.																
REFERENCE	Unclassified.																
AUTHORS	1 (bases 1 to 1091)																
TITLE	Wu,R., You,L. and Soong,T.																
JOURNAL FEATURES	Nucleotide sequence of two circular SSDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus																

[illegible]

79 gggaagtcggtgtttgccaaatafctcggaactcaagccgactggttctacacatgtggt 138

139 ggaaccgaaagacgtattgtaccagtacatcgaggacccaaaacgaaatttaattcctc 198

199 gatgtaccacagtgtaatttagagtattaaataatgccctgttagaatggtttaagaac 258 QY
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
724 GATATCCCGCAGAGTCATTGAGATTCTCAATATGGCGTTATAGACAAATTTAGAAT 783 Db

Db 784 AGAGTTTATTAATPACAAAATACGAACCATGTGTGATTAGAAAAAGATGGACAAAATGTC 843

Db 844 CATGTAATTGTTATGGCAATGTGTTGCCGATTAATGTAATAATTCAGACAGATAGAATA 903

Db 904 AAAATAAATTAATGTTGA 921

LOCUS	1111 bp	DNA	circular	VRL	01-FEB-1995
BBU12586					
BBU12586					

ACCESSION U12586
VERSION U12586.1 GI:542390

ORGANISM Banana bunchy top virus
Viruses; ssDNA viruses; Circoviridae.

TITLE Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related virus

2 (bases 1 to 111)
Wu, R.-Y.
AUTHORS

Division, Development Center for Biotechnology, 81 Chang Hsing Street, Taipei, 10671, Taiwan, ROC

```

source
organism="Banana bunchy top virus"
/db_xref="taxon:12585"

```

Unpublished
2 (bases 1 to 1127)
Wu, R.-Y.
Direct Submission
Submitted (25-JUL-1994) Rey-Yuh Wu, Agricultural Biotechnology
Division, Development Center for Biotechnology, 81 Chang Hsing
Street, Taipei, 10671, Taiwan, ROC
Location/Qualifiers
1. .1127
/organism="Banana bunchy top virus"
/db_xref="taxon:12585"
/clone="S15 and 4-2"
1. .8
15. .49
79. .933
/note="ORF V1"
/codon_start=1
/protein_id="AAA61877.1"
/db_xref="GI:642394"
/translation="MSSFKWCFTLNYSSAEREDFLALLKEELNVAVVGDEVPASS
QKHQQYLSLKKSTKLGKKYSSRAHWEARGSDPDNAKYSCKETLILELGGPASQ
GNNRKLSEWSPSPMRLEQDEIYHYSVKLAKFKKEFVHPCLDRPWQIOLTEA
IDEPDDRTIEWYCPNGNGKSTIAKSLMKDWFYTRGCKENILFYSVYDEGSEKHI
VDFIPRCNQYLNVDVTEALKDRVIESTKYKPKLVELNNIHIVMANFMEPCCKIE
DRKIITYC"
complement(497. .643)
/note="ORF C1"
/codon_start=1
/protein_id="AAA61878.1"
/db_xref="GI:642395"
/translation="MTSHMLISPHSDHRRPRSFQHLVPRQLPPSIESAMVYRGKDE
RIPP"
BASE COUNT 353 a 222 c 260 g 292 t
ORIGIN

Query Match 24.4%; Score 102.8; DB 58; Length 1127;
Best Local Similarity 56.2%; Pred. No. 5.2e-19;
Matches 235; Conservative 0; Mismatches 177; Indels 6; Gaps 2;

Qy 2 ttgaagtgcgtctgcgatcgagagcagcgagcagatcgacaaatctctggtatgag 61
Db 542 TTCAATTGACGGAGCAATTGACGAGGAACACAGATGATCGAAGCAAGTCTTCTGGTCTATG 601
Qy 62 gacgagcggaggagacgggaagtccgtgtttgccaaatatctcgactcaagccgact 121
Db 602 GTCCGAATGTGATGAGGGCAATCAACATATCGGAGTCATT---AATGAAGAAGACT 658
Qy 122 ggtctcacatgctggtgacacagaaagadcatattgtaccagtcacat---cgaggacc 178
Db 659 GGTTCATACACAGAGGTGGGAAGAACGAGAACACTACTGTCTCTACGTGGACGAAGGAT 718
Qy 179 caaaacgaaattcaatctctcgatgtaccagggtgtaatttagatttaaattatgcgcc 238
Db 719 CTGAGAAGCATATGTATTGTGATTCTCCGCGTGAATCAGGATTATTTAAATATTATGATG 778
Qy 239 tgttagaatgtttaagaacaggggcattcagttcggacaaatacgaaccccttagttatc 298
Db 779 TTATAGAGCATTTAAAGGATAGGTCATAGAGAGTACTAAATATATAACCTATTAACTTAG 838
Qy 299 ttgggttcgaacctgcatgtactcgtatttgcgaatgctcgtgattatttgaaaaa 358
Db 839 TTGAATTGAATAATATACATGTAAATTTGTCATGGCTAAATTCATGGCAGAAATTTCTGTAAA 898
Qy 359 tcagcgggacagaataaaactgtggaattattaaagtatgtcatctaaattcac 416
Db 899 TCTCCGAAGCATAGATTAAGATTATTATTGTGTAATAACACGCTATGACAAATCGTAC 956

RESULT 10
FBF132185
LOCUS
DEFINITION
FBE132185 1015 bp DNA circular VRL 16-SEP-2000
faba bean necrotic yellows virus C7-Eg gene, isolate Egyptian

123	gtttacacatgtgtgtggaaccgaagagcgtattgtaccagatcacatcgagaccacaaa	182
635	GTGGTATACAGCTGGAGGAAAGCAGACATATGCTTTACAGCTATTCATTTGACCCAC	694
183	acgaatttaactctcatgtaccaggtgt---aatttagagattttaattatgacct	239
695	ATGCCACCTATGCAATGATATCCCATGATGACAAGAGAATATATTAATTAATTTCTGT	754
240	gttaagattgtttaagacagggcattcagttcggacaaatacgaaccccttagttatct	299
755	AATGACAGATTAAATAGAGTTATTATTAACAAATATGAACCTTGTTACTATTAG	814
300	tggtgtg---accatgtgcattctactgtatttgcacaaatgcctgcctgattttgaa	356
815	AGACGATGTCATATATGTTCAATGTTATTTTGTGTAATTTTACCTGATGTAACAAG	874
357	aatcagagggagcagaataaaactgtggaataatttaagt	396
875	AATCAGTGAAGATAGATAAAATAAATAAATGTTGAAT	914
RESULT 11		
LOCUS	AF216221	1109 bp DNA VRL 08-MAR-2001
DEFINITION	Banana bunchy top virus satellite S1 replication initiation protein (ORF V1) gene, complete cds.	
ACCESSION	AF216221	
VERSION	AF216221.1	GI:12004325
KEYWORDS	Banana bunchy top virus.	
SOURCE	Banana bunchy top virus.	
ORGANISM	Viruses; ssDNA viruses; Circoviridae.	
REFERENCE	1 (bases 1 to 1109)	
AUTHORS	Horser,C., Harding,R. and Dale,J.	
TITLE	Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein	
JOURNAL	J. Gen. Virol. 82 (Pt 2), 459-464 (2001)	
MEDLINE	21102983	
PUBMED	11161286	
REFERENCE	2 (bases 1 to 1109)	
AUTHORS	Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.	
TITLE	Additional rep-encoding DNAs associated with banana bunchy top virus	
JOURNAL	Arch. Virol. 146 (1), 71-86 (2001)	
REFERENCE	3 (bases 1 to 1109)	
AUTHORS	Horser,C.L., Karan,M., Harding,R.M. and Dale,J.L.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-DEC-1999) School of Life Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia	
FEATURES		
source	1. .1109	
	/organism="Banana bunchy top virus"	
	/isolate="Taiwan"	
	/db_xref="taxon:12585"	
	/note="satellite S1"	
stem_loop	3. .33	
	/note="putative sequence"	
mRNA	63. .917	
	/gene="ORF V1"	
gene	63. .917	
	/gene="ORF V1"	
CDS	63. .917	
	/gene="ORF V1"	
	/codon_start=1	
	/product="replication initiation protein"	
	/protein_id="AAG44003.1"	
	/db_xref="GI:12004326"	
	/translation="MSSPKCFTLNYSAAERDFALLKPEELNVAVGVDEVPSSG QHLOQYSLKSIKGLKKYSSRAHWARGSDENAKYCSKETLLELGFPSAQ GSNRKLSEWSRSRPEIHYTSVKLLKPKKEFVHPCIDLRDWOQLTEA IDEEPDRSIIWYGFNGNEGKSTYAKSLMKKDWFTYTRGKKENILFSYVDESEKHI	

VFDPICPNQDYLNVDTIEALKDRVISTKYKPKIKLVELINLHVIVMANFPECKLISE									
DRIKIYYC"									
BASE COUNT	346 a	213 c	257 g	293 t					
ORIGIN									
Query Match	23.3%; Score 98.4; DB 58; Length.1109;								
Best Local Similarity	56.6%; Pred. No. 9.7e-18;								
Matches	224;	Conservative	0;	Mismatches	166;	Indels	6;	Gaps	2;
QY	2	tlgaagtgctgctgcgattcgagagccagcagcagatcgcacaaatcctctggatgctg	61						
Db	526	TTCAATTGACGGAGGCAATTTGACGAGGAACCGATGATCGAAGCATCATCTGGGCTATG	585						
QY	62	gacgagcagggagacgggaagtcggtgtttgccaatatctcggactcaagcccgact	121						
Db	586	GTCCGNAATGCTAATGAGGGGAATCAACATATGCGAAGTCATT---AATGAAGAAGGAT	642						
QY	122	ggtttcacacatggtgtggaaccagaaaggacgctattgtaccagtcacat---cgaggacc	178						
Db	643	GGTTCTACACCAGAGGTTGGGAAGAGGAGACATACATGTTCTTCTTACGTGGACGAAGAT	702						
QY	179	caaaacgaaatttaattcctcgatgtaccagcaggtgtaatttagagtatttaattatgccc	238						
Db	703	CTGAGAAGCATATGTGATTTGATATTCCTCGCTGTAATCAGGATTTTAAATTTATGATG	762						
QY	239	tgttagaattgttgaagacagggcattcagttcggacaaatacgaaccccttagttatc	298						
Db	763	TTATAGAGGCATTAAGGATAGGTGATAGAGAGTACTAAATATAAACCTATTAAAGTTAG	822						
QY	299	tgggttcaccatgtgcgtactgtactgtatttgcgaattctcgcctgattattgaaaa	358						
Db	823	TTGAATTGATTAAATATACATGTAAATTGTCATGGCTAAATTCATGCGCAGAAATCTCTG	882						
QY	359	tcagcagggacagaaataaaactgtggaalatttaaa	394						
Db	883	TCCTCGAAGATAGAAATAAAGATTATTATTGTTAAA	918						
RESULT 12									
LOCUS	NYV5964	1014 bp DNA circular VRL	13-JAN-1999						
DEFINITION	faba bean necrotic yellows virus C7 gene.								
ACCESSION	AJ005964								
VERSION	AJ005964.1	GI:3550512							
KEYWORDS	C7 gene; component 7; putative; rep protein; replication associated protein.								
SOURCE	faba bean necrotic yellows virus.								
ORGANISM	Viruses; ssDNA viruses; Nanovirus.								
REFERENCE	1 (bases 1 to 1014)								
AUTHORS	Katul,L.								
TITLE	Direct Submission								
JOURNAL	Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY								
REFERENCE	2 (bases 1 to 1014)								
AUTHORS	Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.								
TITLE	Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome								
JOURNAL	J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)								
MEDLINE	99094637								
FEATURES									
source	1. .1014								
	/organism="faba bean necrotic yellows virus"								
	/virion								
	/isolate="SV292-88"								
	/db_xref="taxon:59817"								
stem_loop	1. .33								
gene	60. .1013								
	/gene="C7"								
CDS	60. .911								

VFEDIPRCNQDLYNDVIEALKDRVLESTKYKPKIKLVLELINIHVIMANFPECKISE			
DRIKIYC			
BASE COUNT	346 a	213 c	257 g
ORIGIN	293 t		
Query Match	23.3%; Score 98.4; DB 58; Length 1109;		
Best Local Similarity	56.6%; Pred. No. 9.7e-18;		
Matches 224; Conservative	0; Mismatches 166; Indels 6; Gaps 2;		
QY	2	ttgaagtcgtctgcgtcggagagaccgagcagatcgacacatcctctggatatcg	61
Db	526	TTCATTTACGGAGGCAATTCACGAGGAACCCGATGATCGAAGCATCATCTGGGCTATG	585
QY	62	gacgagagcagagagcggaagtcggtttgcacaaatatctcggactcaagccgact	121
Db	586	GTCCGAATGCTTAATGAGGGGAATCAACATATCGGAAGTCATT---AATGAAGAAGGACT	642
QY	122	ggtttcacacatgtggttggaaccgaaagcagctattgtaccagtcacat---cgaggacc	178
Db	643	GGTTCTACACCAGAGGTGGGAAGAAGGAGAACATACGTCTCTTACGTGGACGAAGGAT	702
QY	179	caaacgaaatttaactcctcgatgtaccaggtgttaatttagagattttaaatatgccc	238
Db	703	CTGAGAAGCATATTTGATTTGATATTCCTCGCTGAATCAGGATTTATTAATATGATG	762
QY	239	tgtagaatgtttaagaaacagggcattcagttcggacaaatacagaccccttagttatc	298
Db	763	TTATAGAGCCATTAAAGGATAGGCTGATAGAGAGTACTAAATATAAACCTATTAACTAG	822
QY	299	ttgggttcgaccatgtgcattctactcgtatttgcacaaatcctgcctgcctgatttataaa	358
Db	823	TTGAATTGATTAATATACATGTAATTCATGCTAATTTTCATGCCAGAAATTCGTAAAA	882
QY	359	tcagcagggacagataaaacttggaalattttaa	394
Db	883	TCTCCGAAGATAGATAAAGATTTATTTATTTCTTAAA	918
RESULT 12			
LOCUS	NYV5964	1014 bp DNA	circular VRL 13-JAN-1999
DEFINITION	faba bean necrotic yellows virus C7 gene.		
ACCESSION	AJ005964		
VERSION	AJ005964.1 GI:3550512		
KEYWORDS	C7 gene; component 7; putative; rep protein; replication associated protein.		
SOURCE	faba bean necrotic yellows virus.		
ORGANISM	faba bean necrotic yellows virus		
REFERENCE	1 (bases 1 to 1014)		
AUTHORS	Katul,L.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAY-1998) Katul L., Institute fuer Biochemie und Pflanzenvirologie, Biologische Bundesanstalt fuer Land- und Forstwirtschaft, Messeweg 11 - 12, 38104 Braunschweig, GERMANY		
REFERENCE	2 (bases 1 to 1014)		
AUTHORS	Katul,L., Timchenko,T., Gronenborn,B. and Vetten,H.J.		
TITLE	Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome		
JOURNAL	J. Gen. Virol. 79 (Pt 12), 3101-3109 (1998)		
MEDLINE	99094637		
FEATURES	Location/Qualifiers		
source	1. .1014		
	/organism="faba bean necrotic yellows virus"		
	/viroion		
	/isolate="SV292-88"		
	/db_xref="taxon:59817"		
	1. .33		
	60. .1013		
	/gene="C7"		
	60. .911		
stem_loop			
gene			
CDS			

/gene="C7"
/function="putative replication associated (rep) protein"
/codon_start=1
/product="component 7"
/protein_id="CAA06787.1"
/db_xref="GI:3550513"
/db_xref="SPPREMBL:O91250"
/translation="MPSIRATHWCFTLNPSGSIPEINWTADVOYSIWQHERVGHDLQ
GYIOMKKHVSLLKKMLLPGLAHLMAKPKKAIKYEYCKKSAIAGPWFYTWISGSH
KRLMERFDEPMLEDPGLYRRLCSRVOMKKIRESCTWNFLRPMQDELKLTIEQ
EPDRLTIWYPHGGEKSAFAKYLTKESWMTAGGKATDMLYSYSLDPTCHVICID
IPCTREEYINYSVIEQIKNRVLIINTKEPCTIRDDGHNVHVFPCNELPDVTRISED
RIKILINC"
polyA_signal 899..904
/gene="C7"
TATA_signal 1008..1013
/gene="C7"

BASE COUNT 341 a 173 c 229 g 271 t
ORIGIN

Query Match 22.9%; Score 96.8; DB 59; Length 1014;
Best Local Similarity 55.6%; Pred. No. 2.8e-17;
Matches 229; Conservative 0; Mismatches 177; Indels 6; Gaps 2;

QY 3 tgaagtgtctgtcgatcgagagccagcgagcgatcgacacgaatcctctcgatgacg 62
DB 515 TGAATTGTTGAAGACGATTGAACAGGACACGATGACCGTCAATATTGGGCTATGG 574
QY 63 acgagacgagagagaggaagtcctgtttcccaaatatctcgacacgaagccgactg 122
DB 575 CCCACATGGGAGGAGAGGTAAGACGCGTTTCGGAATACTTAACTTCAAGAGGTTG 634
QY 123 gttctacacatgtgtggaacagagacgtattgtaccagtcacatcgagagcccaaa 182
DB 635 GTGCTATACAGCTGGAGGAGGACACATATGCTTTACACCTATTCTTGGACCCAC 694
QY 183 acgaaatttaactctcgatgacccagtggt---aattagatatttaattatgacct 239
DB 695 ATCCACGATGATGATTGATTTCCAGATGTACAAAGAGAAATATATAATTACTCTGT 754
QY 240 gttagaatgtttaagacagagcattcagttcggacaaatacgaaccccttagtatct 299
DB 755 AATTGAACAGATTAAATAGAGTATTATTAAATACAAATATGAACCTTGTACTATTAG 814
QY 300 tgggttcg---accatgtcgtactcgtatttgcgaatgctcgtcgtattttaa 356
DB 815 AGACGATGTCATATATGTCATGTAATTGTTGTAATTTTACCGATGTAAACAG 874
QY 357 aatcagcagggacagaaataaactgtgaattttaaagtatgtgcatcta 408
DB 875 GATCAGTGAAGATAGANTAAATAAATTTGTTGAATTTGTTTAAAGTA 926

RESULT 13
AR063451 1106 bp DNA PAT 29-SEP-1999
LOCUS
DEFINITION Sequence 3 from patent US 5846705.
ACCESSION AR063451
VERSION AR063451.1 GI:5992759
SOURCE
KEYWORDS Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1106)
AUTHORS Wu, R., You, L., and Soong, T.
TITLE Nucleotide sequence of two circular ssDNA associated with banana bunchy top virus and method for detection of banana bunchy top virus
JOURNAL Patent: US 5846705-A 3 08-DEC-1998;
FEATURES Location/Qualifiers
1..1106
/organism="unknown"

BASE COUNT 335 a 225 c 257 g 289 t
ORIGIN

Query Match 22.9%; Score 96.8; DB 9; Length 1106;
Best Local Similarity 52.8%; Pred. No. 2.8e-17;
Matches 209; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 2 ttgaagtgtctgtcgatcgagagccagcgagacatcgacacacatcctctgatatcg 61
DB 531 TTCAATTGACGGAGCAATTTGACGAGGAAACCCGATGATGAGCAATCATCTGGTCTATG 590
QY 62 gacgagacgagagagcggaagtcctgtgtttgccaaatatctcgcgactcaagcccgact 121
DB 591 GTCCTTATGTAATGAGGGTAAATCAACATATCGGAAGTCACTAATCAAGAAGGATGGT 550
QY 122 gttctacacatgtgtggaacagagaaagagcgtattgtaccagtcacatcgagagcccaa 181
DB 651 TCTACACCGGGTGGGAGAGAGGAGAAATATCTTATCTCTATGTGGACGAGGATCTG 710
QY 182 aacgaaatttaactctcgatgtaccaggtgtaatttagagattttaaattatgacctgt 241
DB 711 ACAAGCATATAGTATTGTATTTCTCTGTTGTAATCAGGATATTTTAAATTTATGATGTA 770
QY 242 tagaagtgtttaagaacagggcattcagttcggacaaatacgaaccccttagttatcttg 301
DB 771 TAGAGGCATTTAAAGGATAGGGTTATAGACAGTACTAAATACAAACCCATAAAGATAGTTG 830
QY 302 ggttcgacacatgcatgtaactcgtatttgcgaatgctcgtcgtcgtatttgaataatca 361
DB 831 AATTAGGTAAATACATGTAATCGTCATGCGAATTTTCATGCTGACTTCTCTAAATCT 890
QY 362 gcaggacagaaataaaactgtggaattttaagta 397
DB 891 CCGAGATCGAATAAAAAATCAUTTTATGCTGAAGAA 926

RESULT 14
BYTV1 1106 bp DNA VRL 30-OCT-1994
LOCUS
DEFINITION Banana bunchy top virus (BBTV DNA I) V1 and CI-C3 genes, complete cds's.
ACCESSION L321166
VERSION L321166.1 GI:520786
KEYWORDS stem loop.
SOURCE Banana bunchy top virus DNA.
ORGANISM Banana bunchy top virus
REFERENCE 1 (bases 1 to 1106)
AUTHORS Wu, R.-Y., You, L.-R., and Soong, T.-S.
TITLE Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus
JOURNAL Unpublished (1994)
FEATURES Location/Qualifiers
1..1106
/organism="Banana bunchy top virus"
/db_xref="taxon:12585"
TATA_signal 1..7
stem_loop 9..37
gene 62..922
CDS 62..922
/gene="V1"
/gene="V1"
/codon_start=1
/protein_id="AAA51422.1"
/db_xref="GI:520787"
/translation="MSSPSLKWCFITLNYSSAAERFSLKKEEDVHVAVGVDEVAPA
TCQHLQGLVLSLKRGLRGLKKYGSRAHWEIARGTDEENSKYCSKETLLLELGFVP
VNSGNKKIISEMVARSPDRMKTEQPEIFHYOSVANKLKFKEEFVHPCLDPRWQILT
EADIEPDDRSIIWYGPYGNEGKSTYAKSLIKDWFTYRGCKENILFLSYVDGSDK
HIVFDIPRCNDYLIYDVEALKDRVIESTKYKPIKIVELGKIHVIMANPMPDFCKI
SEDRKIILYC"
complement(280..426)
gene

CDS	/gene="C3" complement(280..426) /gene="C3" /codon_start=1 /protein_id="AAA51423.1" /db_xref="GI:520788" /translation="MRSGERATISIDFLLEPLTTNPNSRIRVSFEQYFFESSVPL ALSO" complement(636..779) /gene="C1" complement(636..779) /gene="C1" /codon_start=1 /protein_id="AAA51424.1" /db_xref="GI:520789" /translation="MPLLHNNINPDYNEEYOILYACQLRPHRRIYSPSSHPWCRT NPS" complement(760..912) /gene="C2" complement(760..912) /gene="C2" /codon_start=1 /protein_id="AAA51425.1" /db_xref="GI:520790" /translation="MIFRSSEIILOKSMGFAMTITCILPNSIFMGLYLVLSTLSL NASITS"
BASE COUNT	335 a 225 c 257 g 289 t
ORIGIN	
Query Match	22.9%; Score 96.8; DB 58; Length 1106;
Best Local Similarity	52.8%; Pred.No.2.8e-17;
Matches	209; Conservative 0; Mismatches 187; Indels 0; Gaps 0;
DQ	2 ttgaagtgtctgcgatacggagccgacgacatcgcaaatcctctggatatcgy 61
DB	531 TTCAAATTGACGGAGCAATTCACGAGAACCCGATGATCGAAGCATCATCTGGTCTATG 590
DQ	62 gacgagacgaggagacggagaatccggttttgccaaatatctcggaactcaagcccga 121
DB	591 GTCTCTTATGTGAATGAGGGTAATCAACATATGCGAAGTCACTAATCAAGAAGATTGGT 650
DQ	122 ggcttcacatgtggtggaccgaaagacgtattgaccagtacatcgcgagagccccaa 181
DB	651 TCTACACCAGGGTGGAAGAGAGAAATCTTATTCTCTATGTGGACGAAGATCTG 710
DQ	182 aacgaaaatttaactoctcgatgtacctcaggtgtaatttagagattttaaatatgccctg 241
DB	711 ACAAGCATATAGTATTGTATTCCTCGTTGTAAATCAGGAATATTTAAATTAATGATGTAA 770
DQ	242 tagaatgtttaagaacaggcatcagttcggacaaaaacaccccccttagtattcttg 301
DB	771 TAGAGGCATTAAGCATAGGTTATACAGAGACTATAAATACAACCCATAAAGATAGTTG 830
DQ	302 ggctcgaccatgtcgatgtactcgtatctggccaatgtcctgcctgatttggaaatca 361
DB	831 AATTAGGTAAAAATACATGTAATCCTCATGCCGAATTTTCATGCTGACTTCTGTAAATCT 890
DQ	362 gcaggacacaataaaaactgtggaatttaaagta 397
DB	891 CCGAAGCATGATATAAATCATTTATTGCTCAAGAA 926
RESULT 15	
SCU16735	SCU16735 1017 bp DNA circular VRL 07-JUN-1995
LOCUS	Subterranean clover stunt virus possible replication associated
DEFINITION	protein (SCSV6) gene, complete cds.
ACCESSION	U16735
VERSION	U16735.1 GI:571495
KEYWORDS	. Subterranean clover stunt virus.
SOURCE	Subterranean clover stunt virus
ORGANISM	

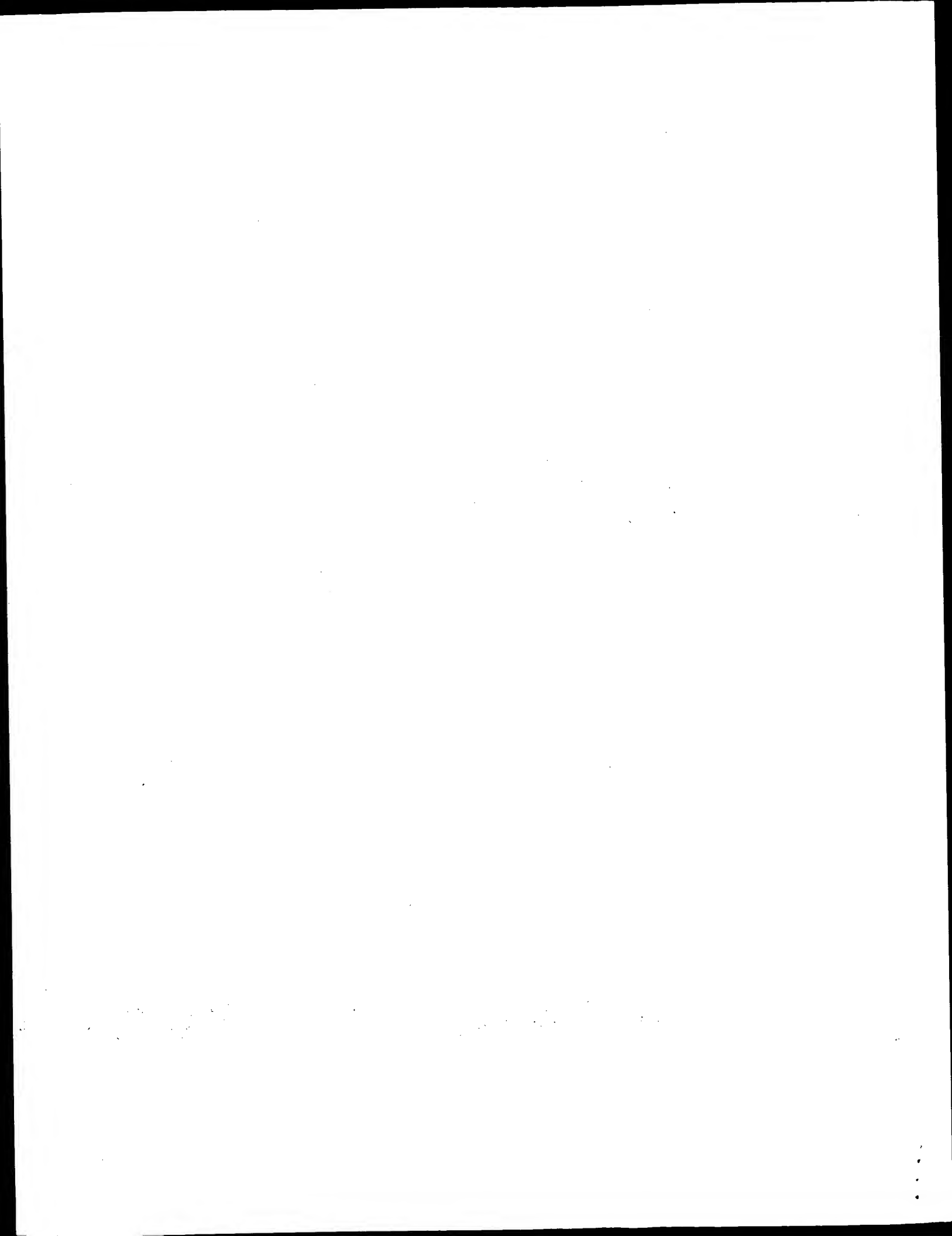
Thu Oct 25 13:08:14 2001

us-09-462-955-1_copy_583_1004.rge

Page 11

Db 873 ACTGAAGATAGATAAAATAATTCGTTGTTGCARAACTCTG 913

Search completed: October 24, 2001, 11:42:35
Job time: 6340 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:08 ; Search time 5701.85 Seconds
(without alignments)
84.096 Million cell updates/sec

Title: US-09-462-955-1-copy_941_971
Perfect score: 31
Sequence: 1 tcagcaggacagagaataaaactgtggaatat 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_om:*
9: gb_ov:*
10: gb_pat1:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vil2:*
60: gb_vil2:*
61: gb_vil2:*
62: gb_vil2:*
63: gb_vil2:*
64: gb_vil2:*
65: gb_vil2:*
66: gb_vil2:*
67: gb_vil2:*
68: gb_vil2:*
69: gb_vil2:*
70: gb_vil2:*
71: gb_vil2:*
72: gb_vil2:*
73: gb_vil2:*
74: gb_vil2:*
75: gb_vil2:*
76: gb_vil2:*
77: gb_vil2:*
78: gb_vil2:*
79: gb_vil2:*
80: gb_vil2:*
81: gb_vil2:*
82: gb_vil2:*
83: gb_vil2:*
84: gb_vil2:*
85: gb_vil2:*
86: gb_vil2:*
87: gb_vil2:*
88: gb_vil2:*
89: gb_vil2:*
90: gb_vil2:*
91: gb_vil2:*
92: gb_vil2:*
93: gb_vil2:*
94: gb_vil2:*
95: gb_vil2:*
96: gb_vil2:*
97: gb_vil2:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	31	100.0	1291	58	CFDCG
2	21.2	68.4	137949	88	AC022137
c 3	21.2	68.4	162107	83	AP002423
4	21.2	68.4	163192	76	AC079751
c 5	21.2	68.4	163192	76	AC079751
6	21.2	68.4	177748	64	AC016185
c 7	21.2	68.4	181041	77	AC087677
8	21.2	68.4	181239	68	AC023983
					M29963 Coconut fol
					AC022137 Homo sapi
					AP002423 Homo sapi
					AC079751 Homo sapi
					AC079751 Homo sapi
					AC016185 Homo sapi
					AC087677 Homo sapi
					AC023983 Homo sapi

TRGVKRLPRAQFAEEDELKLEDPGGYRCVVGASVETWRWAAENPPFPVHNWQ
LEVLNLAIGSPADRRILITCGDGGDKSVFAKYLGLKPDWFTYTCGTRKDLVLYLE
DKRNLILDPNCLNLEYLALLECCKNRAFSSDKYEPVLSYLGEDHVLVLFANVLPD
YLKISDRDKLWNI"
314..775
/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSPVVTNRITVRRNGCFSPSTESRLVLESKGDWPN
DLRLNMSAWKTQADTEDALTELWNCQDGPLKIRSHFHTITGSLKCCILRSORT
IAOSSYADETETSPCLPNISDSPTGSHVVEPFTYCTISRTQNEI"
complement(422..568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFORPILSDPPKLRVORIFGIRLGLPGGVHQPQIVGP
IVAF"
639..797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MTRRRRRRVRCQISRTQARLVLMWNNQKGRIVPVHRGPKTK
FNPRCTGV"
complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="MTHLNIPOFYSVPADFQIIRODICKYEYMHMVEPKITKGFVVF
RTECPVLNTP"
1098..1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGPTIKDSIWIRTNLLCLOCTOPLSTSPISQVSSLEKK
AASLYLPSICFCAIGRLS"
BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN

Query Match 100.0%; Score 31; DB 58; Length 1291;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagcaggagacagaataaaactgtggaatat 31
|||||
Db 941 TCAGCAGGACAGCAATAAACTGTGGAATAT 971

RESULT 2
AC022137 137949 bp DNA PRI 26-NOV-2000
LOCUS Homo sapiens chromosome 19 clone CTD-2224J9, complete sequence.
AC022137
ACCESSION
VERSION AC022137.5 GI:11079415
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 137949)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 137949)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

ALL09748 Homo sapi
AL162851 Homo sapi
AL050303 Homo sapi
AL163204 Homo sapi
AC021220 Homo sapi
AB029530 Mus muscu
E13044 cDNA of mou
E21655 Spermatogen
D78270 Mouse mRNa
AC087067 Rattus no
AL429486 clone BA0
AC026725 Homo sapi
AC079045 Homo sapi
AC026706 Homo sapi
AL137853 Human DNA
AL158012 Human DNA
AL359095 Homo sapi
AC069475 Homo sapi
AC028339 Homo sapi
AL160173 Homo sapi
AL451047 Homo sapi
AL1513175 Homo sapi
AL138778 Homo sapi
AC055798 Homo sapi
AL353714 Homo sapi
AC027429 Homo sapi
AC011454 Homo sapi
AC024511 Homo sapi
AC067904 Homo sapi
AC074173 Mus muscu
AC074144 Mus muscu
AC074144 Mus muscu
AL357892 Human DNA
AL391831 Homo sapi
AL157367 Homo sapi
AL450169 Genomic s
AF279660 Homo sapi

11 67.7 73465 92 HS256M13
21 67.7 152318 79 AL162851
11 67.7 168075 84 HS214G14
21 67.7 340000 92 HS21C004
13 67.1 214348 66 AC021220
14 66.5 1211 94 AB02952S11
15 20.6 4621 10 E13044
16 20.6 4621 10 E21655
17 20.6 4621 94 D78270
18 20.6 178010 77 AC087067
19 65.8 866 53 CNS07508
20 20.4 65.8 101856 70 AC026725
21 20.4 65.8 110620 88 AC079045
22 20.4 65.8 111489 70 AC026706
23 20.4 65.8 113853 89 AL137853
24 20.4 65.8 124474 90 AL158012
25 20.4 65.8 141514 80 AL359095
26 20.4 65.8 153092 88 AC069475
27 20.4 65.8 157069 67 AC028339
28 20.4 65.8 159774 79 AL160173
29 20.4 65.8 161545 81 AL451047
30 20.4 65.8 164782 81 AL1513175
31 20.4 65.8 167060 78 AL138778
32 20.4 65.8 167314 72 AC055798
33 20.4 65.8 167333 79 AL353714
34 20.4 65.8 167685 70 AC027429
35 20.4 65.8 171829 62 AC011454
36 20.4 65.8 173077 68 AC024511
37 20.4 65.8 173409 73 AC067904
38 20.4 65.8 180129 75 AC074173
39 20.4 65.8 184800 75 AC074144
40 20.4 65.8 184800 75 AC074144
41 20.4 65.8 189876 90 AL357892
42 20.4 65.8 200930 81 AL391831
43 20.4 65.8 245465 79 AL157367
44 20.4 65.8 279647 91 CNS07EEW
45 20.4 65.8 347503 89 AF279660

ALIGNMENTS

RESULT 1
CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993
LOCUS Coconut foliar decay virus, complete genome.
DEFINITION M29963
ACCESSION
VERSION M29963.1 GI:323306
KEYWORDS circular; complete genome.
SOURCE Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSUA3A.
ORGANISM Coconut foliar decay virus
Viruses; ssDNA viruses; Nanovirus.
REFERENCE 1 (bases 1 to 1291)
Rohde,W., Randles,J.W., Langridge,P. and Hanold,D.
Nucleotide sequence of a circular single-stranded DNA associated
with coconut foliar decay virus
Virology 176, 648-651 (1990)
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by
W.Rhode, 15-MAR-1989, for release after publication.
FEATURES
source
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
stem_loop
40..70
/note="stem-loop structure"
CDS
103..975
/notes="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSTRRWCFTLNYETEAEANVRRLESNLVYAVIGDEVAPS
TGORHLQGFHLTKTRRLQGLKTVLIGNDRHLEPTGSDQNRDSCSKERYVLEHGVP

```

REFERENCE
AUTHORS   3 (bases 1 to 137949)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
Submitted (03-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS   4 (bases 1 to 137949)
TITLE     DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL   Direct Submission
Submitted (26-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT    On Nov 3, 2000 this sequence version replaced gi:9964763.
Draft sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.4.
STS Content:
SHGC-84347 G60949
SHGC-103885 G57910.

FEATURES             source
   source
1..137949
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2224J9"
BASE COUNT    39207 a 31405 c 30153 g 37184 t
ORIGIN
Query Match      68.4%; Score 21.2; DB 88; Length 137949;
Best Local Similarity 88.5%; Pred. No. 74;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcaqcgagacaaataaaactgtg 26
||||| ||||| ||||| ||||| |||||
Db 16523 TCAGGAGTACAGAAATAAACATGG 16548

RESULT 3
AP002423/c
LOCUS      Homo sapiens chromosome 18 clone RP11-863L8 map 18q11.2, WORKING
DEFINITION DRAFT SEQUENCE, 42 unordered pieces.
ACCESSION  AP002423
VERSION     AP002423.2 GI:8439535
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-863L8.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 162,107 genomic DNA of 18q11.2
Published Only in Database (2000) In press
2 (bases 1 to 162107)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hqp.gsc.riken.go.jp,
Fax:81-42-778-9924)
On Jun 9, 2000 this sequence version replaced gi:8131687.
-----Genome Center
Center code: RIKEN
Web site: http://hqp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information

```

```

Center project name: HumDraft18
Center clone name: RP11-863L8
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 143757 bases at least Q40
Consensus quality: 152091 bases at least Q30
Consensus quality: 155858 bases at least Q20
Insert size: 158007; sum-of-contigs
Quality coverage: 4.32x in Q20 bases; sum-of-contigs
-----

```

NOTE: This is a 'working draft' sequence. It currently consists of 42 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 12846 contig of 12846 bp in length
12947 23222 contig of 10276 bp in length
23323 31114 contig of 7792 bp in length
31215 38997 contig of 7783 bp in length
39098 46783 contig of 7686 bp in length
46884 54555 contig of 7772 bp in length
54756 59749 contig of 4994 bp in length
59850 65160 contig of 5311 bp in length
65261 71291 contig of 6031 bp in length
71392 75257 contig of 3866 bp in length
75358 79204 contig of 3847 bp in length
79305 83972 contig of 4668 bp in length
84073 88891 contig of 4819 bp in length
88992 93736 contig of 4745 bp in length
93837 98281 contig of 4445 bp in length
98382 102273 contig of 3892 bp in length
102374 106147 contig of 3774 bp in length
106248 110389 contig of 4142 bp in length
110490 112888 contig of 2399 bp in length
112989 115657 contig of 2669 bp in length
115758 119437 contig of 3680 bp in length
119538 123011 contig of 3474 bp in length
123112 125411 contig of 2300 bp in length
125512 128514 contig of 3003 bp in length
128615 131919 contig of 3305 bp in length
132020 133329 contig of 1310 bp in length
133430 136197 contig of 2768 bp in length
136298 138656 contig of 2359 bp in length
138757 140802 contig of 2046 bp in length
140903 142951 contig of 2049 bp in length
143052 145259 contig of 2208 bp in length
143360 147234 contig of 1875 bp in length
147335 149339 contig of 2005 bp in length
149440 150983 contig of 1544 bp in length
151084 152215 contig of 1132 bp in length
152316 153476 contig of 1161 bp in length
153577 155210 contig of 1634 bp in length
155311 156633 contig of 1323 bp in length
156734 157980 contig of 1247 bp in length
158081 159424 contig of 1344 bp in length
159525 160646 contig of 1122 bp in length
160747 162107 contig of 1361 bp in length

```

NOTE: This is a 'working draft' sequence. It currently consists of 42 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 12846: contig of 12846 bp in length
12947 12946: gap of 100 bp
23222: contig of 10276 bp in length

```

Thu Oct 25 13:08:32 2001

* 23223 23322: gap of 100 bp
* 23323 31114: contig of 7792 bp in length
* 31115 31214: gap of 100 bp
* 31215 38997: contig of 7783 bp in length
* 38998 39097: gap of 100 bp
* 39098 46783: contig of 7686 bp in length
* 46784 46883: gap of 100 bp
* 46884 54655: contig of 7772 bp in length
* 54656 54755: gap of 100 bp
* 54756 59749: contig of 4994 bp in length
* 59750 59849: gap of 100 bp
* 59850 63160: contig of 5311 bp in length
* 63161 65260: gap of 100 bp
* 65261 71291: contig of 6031 bp in length
* 71292 71391: gap of 100 bp
* 71392 75257: contig of 3866 bp in length
* 75258 75357: gap of 100 bp
* 75358 79204: contig of 3847 bp in length
* 79205 79304: gap of 100 bp
* 79305 83972: contig of 4668 bp in length
* 83973 84072: gap of 100 bp
* 84073 88891: contig of 4819 bp in length
* 88892 89991: gap of 100 bp
* 89992 93736: contig of 4745 bp in length
* 93737 93836: gap of 100 bp
* 93837 98281: contig of 4445 bp in length
* 98282 98381: gap of 100 bp
* 98382 102273: contig of 3892 bp in length
* 102274 102373: gap of 100 bp
* 102374 106147: contig of 3774 bp in length
* 106148 106247: gap of 100 bp
* 106248 110389: contig of 4142 bp in length
* 110390 110489: gap of 100 bp
* 110490 112888: contig of 2399 bp in length
* 112889 112988: gap of 100 bp
* 112989 115657: contig of 2669 bp in length
* 115658 115757: gap of 100 bp
* 115758 119437: contig of 3680 bp in length
* 119438 119537: gap of 100 bp
* 119538 123011: contig of 3474 bp in length
* 123012 123111: gap of 100 bp
* 123112 125411: contig of 2300 bp in length
* 125412 125511: gap of 100 bp
* 125512 128514: contig of 3003 bp in length
* 128515 128614: gap of 100 bp
* 128615 131919: contig of 3305 bp in length
* 131920 132019: gap of 100 bp
* 132020 133329: contig of 1310 bp in length
* 133330 133429: gap of 100 bp
* 133430 136197: contig of 2768 bp in length
* 136198 136297: gap of 100 bp
* 136298 138656: contig of 2359 bp in length
* 138657 138756: gap of 100 bp
* 138757 140802: contig of 2046 bp in length
* 140803 140902: gap of 100 bp
* 140903 142951: contig of 2049 bp in length
* 142952 143051: gap of 100 bp
* 143052 145259: contig of 2208 bp in length
* 145260 145359: gap of 100 bp
* 145360 147234: contig of 1875 bp in length
* 147235 147334: gap of 100 bp
* 147335 149339: contig of 2005 bp in length
* 149340 149439: gap of 100 bp
* 149440 150983: contig of 1544 bp in length
* 150984 151083: gap of 100 bp
* 151084 152215: contig of 1132 bp in length
* 152216 152315: gap of 100 bp
* 152316 153476: contig of 1161 bp in length
* 153477 153576: gap of 100 bp
* 153577 155210: contig of 1634 bp in length
* 155211 155310: gap of 100 bp
* 155311 156633: contig of 1323 bp in length
* 156634 156733: gap of 100 bp

* 156734 157980: contig of 1247 bp in length
* 157981 158080: gap of 100 bp
* 158081 159424: contig of 1344 bp in length
* 159425 159524: gap of 100 bp
* 159525 160646: contig of 1122 bp in length
* 160647 160746: gap of 100 bp
* 160747 162107: contig of 1361 bp in length.

FEATURES

source
1..162107
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q11.2"
/clone="RP11-863L8"
1..12846
/note="assembly_fragment"
12947..23222
/note="assembly_fragment"
23323..31114
/note="assembly_fragment clone_end:T7 vector_side:right"
31215..38997
/note="assembly_fragment"

misc_feature
1..12846
/note="assembly_fragment"
12947..23222
/note="assembly_fragment"
23323..31114
/note="assembly_fragment clone_end:T7 vector_side:right"
31215..38997
/note="assembly_fragment"

Query Match . 68.4%; Score 21.2; DB 83; Length 162107;

Best Local Similarity 88.5%; Pred. No. 75;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 caggagacagaataaactgtggaata 30

Db 72809 CAGAGACAGATAAATACTGCAGATA 72784

RESULT 4

AC079751 163192 bp DNA HTG 10-SEP-2000

LOCUS Homo sapiens chromosome 19 clone RP11-61M18, WORKING DRAFT

DEFINITION SEQUENCE, 19 unordered pieces.

AC079751

VERSION AC079751.1 GI:10047939

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 163192)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 163192)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H.NH0061M18

----- Summary Statistics -----

Sequencing vector: M13; 100%

Chemistry: Dye-terminator Big Dye; 0% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 150903 bases at least Q40

Consensus quality: 155361 bases at least Q30

Consensus quality: 157169 bases at least Q20

Insert size: 165000; agarose-pp

Insert size: 161392; sum-of-contigs

Quality coverage: 4.38 in Q20 bases; agarose-pp

Insert size: 161392; sum-of-contigs

Quality coverage: 4.38 in Q20 bases; agarose-fp
Quality coverage: 4.56 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
3640: contig of 3640 bp in length
3641: gap of unknown length
3741: contig of 4347 bp in length
8088: gap of unknown length
8188: contig of 5374 bp in length
13561: gap of unknown length
13562: contig of 5238 bp in length
18899: gap of unknown length
18900: contig of 4803 bp in length
23802: gap of unknown length
23903: contig of 6081 bp in length
29983: gap of unknown length
30083: contig of 6758 bp in length
36841: gap of unknown length
36842: contig of 8463 bp in length
45405: gap of unknown length
45505: contig of 8121 bp in length
53625: gap of unknown length
53725: contig of 8995 bp in length
53726: gap of unknown length
62720: contig of 8995 bp in length
62721: gap of unknown length
62821: contig of 9943 bp in length
72764: gap of unknown length
72864: contig of 19901 bp in length
92765: gap of unknown length
92766: contig of 21615 bp in length
92865: gap of unknown length
114480: contig of 21615 bp in length
114580: gap of unknown length
149140: contig of 34561 bp in length
149141: gap of unknown length
150751: contig of 1511 bp in length
150851: gap of unknown length
153922: contig of 3071 bp in length
153923: gap of unknown length
154023: contig of 2730 bp in length
156753: gap of unknown length
156754: contig of 2906 bp in length
159759: gap of unknown length
159859: contig of 3334 bp in length.

Location/Qualifiers
1. .163192
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-61M18"
1. .3640
/note="assembly_name:Contig10"
3741. .8087
/note="assembly_name:Contig11"
8188. .13561
/note="assembly_name:Contig12"
13662. .18899
/note="assembly_name:Contig13"
19000. .23802
/note="assembly_name:Contig14"
23903. .29983
/note="assembly_name:Contig15"
30084. .36841
/note="assembly_name:Contig16"
36942. .45404
/note="assembly_name:Contig17"
45505. .53625
/note="assembly_name:Contig18"
53726. .62720

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

/note="assembly_name:Contig19"
62821. .72763
/note="assembly_name:Contig20"
72864. .92764
/note="assembly_name:Contig21"
92865. .114479
/note="assembly_name:Contig22"
114580. .149140
/note="assembly_name:Contig23"
149241. .150751
/note="assembly_name:Contig5"
150852. .153922
/note="assembly_name:Contig6"
154023. .156752
/note="assembly_name:Contig7"
156853. .159758
/note="assembly_name:Contig8"
vector_side:left
159859. .163192
/note="assembly_name:Contig9"
45259 a 35953 c 35368 g 44806 t 1806 others
ORIGIN

Query Match 68.4%; Score 21.2; DB 76; Length 163192;
Best Local Similarity 88.5%; Pred. NO. 75;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcagcaggacagataaaactgtgg 26
||||| ||| ||||| ||||| ||||| |||||
Db 71667 TCAGGAGGTACAGATAAAACAGTGG 71642

RESULT 6
AC016185 177748 bp DNA HTG 22-NOV-2000
LOCUS Homo sapiens chromosome 18 clone RP11-652111 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 22 unordered pieces.
AC016185
AC016185.4 GI:11276249
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 177748)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-652111
Unpublished
2 (bases 1 to 177748)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A., Doyle,M.,
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2000 this sequence version replaced gi:10835345.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4257

Center clone name: 652-I-11

----- Summary Statistics

Sequencing vector: M13; M77815; 1% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 163169 bases at least Q40

Consensus quality: 170756 bases at least Q30

Insert size: 182000; agarose-fp

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

Quality coverage: 4.3 in Q20 ba.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 558: contig of 558 bp in length
559 658: gap of 100 bp
659 1793: contig of 1135 bp in length
1794 1893: gap of 100 bp
1894 3080: contig of 1187 bp in length
3081 3180: gap of 100 bp
3181 5062: contig of 1882 bp in length
5063 5162: gap of 100 bp
5163 7267: contig of 2105 bp in length
7268 7367: gap of 100 bp
7368 10092: contig of 2725 bp in length
10093 10192: gap of 100 bp
10193 14087: contig of 3895 bp in length
14088 14187: gap of 100 bp
14188 18512: contig of 4325 bp in length
18513 18612: gap of 100 bp
18613 23300: contig of 4688 bp in length
23301 23400: gap of 100 bp
23401 28085: contig of 4685 bp in length
28086 28185: gap of 100 bp
28186 32174: contig of 3989 bp in length
32175 32274: gap of 100 bp
32275 61458: contig of 29184 bp in length
61459 61558: gap of 100 bp
61559 68206: contig of 6648 bp in length
68207 68306: gap of 100 bp
68307 74318: contig of 6012 bp in length
74319 74418: gap of 100 bp
74419 82866: contig of 8448 bp in length
82867 82966: gap of 100 bp
82967 94336: contig of 11370 bp in length
94337 94436: gap of 100 bp
94437 107901: contig of 13465 bp in length
107902 108001: gap of 100 bp
108002 121825: contig of 13824 bp in length
121826 121925: gap of 100 bp
121926 138124: contig of 16199 bp in length
138125 138224: gap of 100 bp
138225 154039: contig of 15815 bp in length
154040 154139: gap of 100 bp
154140 172971: contig of 18832 bp in length
172972 173071: gap of 100 bp
173072 177748: contig of 4677 bp in length.
Location/Qualifiers
1..177748
/organism="Homo sapiens"

```

FEATURES
 source

```

/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-652I11"
/clone_lib="RPC1-11 Human Male BAC"
1..558
/notes="assembly_fragment"
vector_side:left"
clone_end:SP6
659..1793
/notes="assembly_fragment"
1894..3080
/notes="assembly_fragment"
3181..5062
/notes="assembly_fragment"
5163..7267
/notes="assembly_fragment"
7368..10092
/notes="assembly_fragment"
10193..14087
/notes="assembly_fragment"
14188..18512
/notes="assembly_fragment"
18613..23300
/notes="assembly_fragment"
23401..28085
/notes="assembly_fragment"
28186..32174
/notes="assembly_fragment"
32275..61458
/notes="assembly_fragment"
61559..68206
/notes="assembly_fragment"
68307..74318
/notes="assembly_fragment"
74419..82866
/notes="assembly_fragment"
82967..94336
/notes="assembly_fragment"
94437..107901
/notes="assembly_fragment"
108002..121825
/notes="assembly_fragment"
121926..138124
/notes="assembly_fragment"
138225..154039
/notes="assembly_fragment"
154140..172971
/notes="assembly_fragment"
173072..177748
/notes="assembly_fragment"
clone_end:P7
vector_side:right"
53384 a 35337 c 35096 g 51826 t 2105 others
ORIGIN

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 53384 a 35337 c 35096 g 51826 t 2105 others
 ORIGIN

Query Match

Best Local Similarity 68.4%; Score 21.2; DB 64; Length 177748;

Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 cagggacagaataaaactgtggaata 30

|||||
 Db 143615 CAGACAGAGATAAACTGCAGATA 143640

RESULT 7

AC087677/c

LOCUS

DEFINITION Homo sapiens chromosome 18 clone RP11-863L8 map 18; WORKING DRAFT

SEQUENCE, 26 unordered pieces.

ACCESSION AC087677

VERSION AC087677.2 GI:13123907

HTG 26-FEB-2001

Thu Oct 25 13:08:32 2001

us-09-462-955-1-copy_941_971.rge

KEYWORDS

HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.

HTG: HTGS_PHASE1; HTGS_DRAFT.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 181041)

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens chromosome 18, clone RP11-863L8

JOURNAL

Unpublished

AUTHORS

2 (bases 1 to 181041)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barina, N., Bastien, V., Boguslavskiy, L., Boukhaltier, B., Brown, A.,
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., Lakocque, K., Lamazares, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienda, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
 Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (15-JAN-2001)

COMMENT

Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 26, 2001 this sequence version replaced gi:12229452.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12275

Center clone name: 863.L.8

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 172058 bases at least Q40

Consensus quality: 175838 bases at least Q30

Consensus quality: 177418 bases at least Q20

Insert size: 170000; agarose-1p

Insert size: 178541; sum-of-contigs

Quality coverage: 5.8 in Q20 bases; agarose-1p

Quality coverage: 5.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 26 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 7738: contig of 7738 bp in length

* 7739 7838: gap of 100 bp

* 7839 8498: contig of 660 bp in length

* 8499 8598: gap of 100 bp

* 8599 9231: contig of 633 bp in length

* 9232 9331: gap of 100 bp

* 9332 9879: contig of 548 bp in length

* 9880 9979: gap of 100 bp
 * 9980 10601: contig of 622 bp in length
 * 10602 10701: gap of 100 bp
 * 10702 12024: contig of 1323 bp in length
 * 12025 12124: gap of 100 bp
 * 12125 12920: contig of 796 bp in length
 * 12921 13020: gap of 100 bp
 * 13021 13844: contig of 824 bp in length
 * 13845 13944: gap of 100 bp
 * 13945 15024: contig of 1080 bp in length
 * 15025 15124: gap of 100 bp
 * 15125 38235: contig of 2311 bp in length
 * 38236 38335: gap of 100 bp
 * 38336 40005: contig of 1670 bp in length
 * 40006 40105: gap of 100 bp
 * 40106 42557: contig of 2452 bp in length
 * 42558 42657: gap of 100 bp
 * 42658 46488: contig of 3831 bp in length
 * 46489 46588: gap of 100 bp
 * 46589 52126: contig of 5538 bp in length
 * 52127 52226: gap of 100 bp
 * 52227 56744: contig of 4518 bp in length
 * 56745 56844: gap of 100 bp
 * 56845 62088: contig of 5244 bp in length
 * 62089 62188: gap of 100 bp
 * 62189 67648: contig of 5460 bp in length
 * 67649 67748: gap of 100 bp
 * 67749 74574: contig of 6826 bp in length
 * 74575 74674: gap of 100 bp
 * 74675 83911: contig of 9237 bp in length
 * 83912 84011: gap of 100 bp
 * 84012 96752: contig of 12741 bp in length
 * 96753 96852: gap of 100 bp
 * 96853 107204: contig of 10352 bp in length
 * 107205 107304: gap of 100 bp
 * 107305 123659: contig of 16355 bp in length
 * 123660 123759: gap of 100 bp
 * 123760 139462: contig of 15703 bp in length
 * 139463 139562: gap of 100 bp
 * 139563 160301: contig of 20739 bp in length
 * 160302 160401: gap of 100 bp
 * 160402 180220: contig of 19819 bp in length
 * 180221 180320: gap of 100 bp
 * 180321 181041: contig of 721 bp in length.

FEATURES
 source
 1. .181041
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone_lib="RP11-863L8"
 /clone="RP11-863L8"
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 7839..8498
 /note="assembly_fragment"
 8599..9231
 /note="assembly_fragment"
 9332..9879
 /note="assembly_fragment"
 9980..10601
 /note="assembly_fragment"
 10702..12024
 /note="assembly_fragment"
 12125..12920
 /note="assembly_fragment"
 13021..13844
 /note="assembly_fragment"
 13945..15024
 /note="assembly_fragment"
 15125..38235
 /note="assembly_fragment"

misc_feature
 1. .7738
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 7839..8498
 /note="assembly_fragment"
 8599..9231
 /note="assembly_fragment"
 9332..9879
 /note="assembly_fragment"
 9980..10601
 /note="assembly_fragment"
 10702..12024
 /note="assembly_fragment"
 12125..12920
 /note="assembly_fragment"
 13021..13844
 /note="assembly_fragment"
 13945..15024
 /note="assembly_fragment"
 15125..38235
 /note="assembly_fragment"


```

/note="assembly_fragment"
38336. .40005
/note="assembly_fragment"
40106. .42557
/note="assembly_fragment"
42658. .46488
/note="assembly_fragment"
46589. .52126
/note="assembly_fragment"
52227. .56744
/note="assembly_fragment"
56845. .62088
/note="assembly_fragment"
62189. .67648
/note="assembly_fragment"
67749. .74574
/note="assembly_fragment"
74675. .83911
/note="assembly_fragment"
84012. .96752
/note="assembly_fragment"
96853. .107204
/note="assembly_fragment"
107305. .123659
/note="assembly_fragment"
123760. .139462
/note="assembly_fragment"
139563. .160301
/note="assembly_fragment"
160402. .180220
/note="assembly_fragment"
180321. .181041
/note="assembly_fragment
vector_side:right"
clone_end:T7
BASE COUNT 52864 a 37195 c 36606 g 51876 t 2500 others
ORIGIN

Query Match 68.4%; Score 21.2; DB 77; Length 181041;
Best Local Similarity 88.5%; Pred. No. 75;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 caggacagacgaataaaactgtgaata 30
||||| ||||| ||||| ||||| |||||
Db 94662 CAGACAGACGATTAACCTGCAGAATA 94637

RESULT 8
AC023983
LOCUS AC023983 181239 bp DNA HTG 25-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-62116 map 18, WORKING DRAFT
SEQUENCE, 35 unordered pieces.
ACCESSION AC023983
VERSION AC023983.3 GI:8076908
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181239)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-62116
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181239)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baidwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouckgaier,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Menes,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Ollivar,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7212065.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L6883
Center clone name: 621_L_6

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160357 bases at least Q40
Consensus quality: 170823 bases at least Q30
Consensus quality: 174839 bases at least Q20
Insert size: 183000; agarose-fp
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contents

* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1064: contig of 1064 bp in length
* 1065 1164: gap of 100 bp
* 1165 2658: contig of 1494 bp in length
* 2659 2758: gap of 100 bp
* 2759 4357: contig of 1599 bp in length
* 4358 4457: gap of 100 bp
* 4458 5644: contig of 1187 bp in length
* 5645 5744: gap of 100 bp
* 5745 7870: contig of 2126 bp in length
* 7871 9790: gap of 100 bp
* 9791 9676: contig of 1706 bp in length
* 9677 9776: gap of 100 bp
* 9777 11565: contig of 1789 bp in length
* 11566 11665: gap of 100 bp
* 11666 12026: contig of 361 bp in length
* 12027 12126: gap of 100 bp
* 12127 14427: contig of 2301 bp in length
* 14428 14527: gap of 100 bp
* 14528 17145: contig of 2618 bp in length
* 17146 17245: gap of 100 bp
* 17246 20711: contig of 3466 bp in length
* 20712 20811: gap of 100 bp
* 20812 23354: contig of 2543 bp in length
* 23355 23454: gap of 100 bp
* 23455 27289: contig of 3835 bp in length

[illegible]

feature lines that evidence is not experimental.

```

PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+ + + + +
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'AnnoMitter' (Hornischer & Bloecker).
+ + + + +
+ Programs used by 'AnnoMitter':
+ + + + +
> GeneFinder (Green), Vers. 084
> Organism: human
> Genscan (Burge & Karlin), Vers. 1.0
> Used matrix: vertebrate; Minimum score: 0
> Grail (Xu et al.), Vers. 1.3
> Organism: human
> Mzef (Zhang)
> Skolnick
> Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
  & Skolnick)
> Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
  BLASTN 2.0.9 (Altschul et al.)
  * Database(s): * RepBase (human), released 22-DEC-1995
  * RepBase (mammal), released 22-DEC-1995
  * Minimum identity: 70 %; Minimum score: 60;
  > 'ESTs': BLASTN 2.0.9 (Altschul et al.)
  * Database(s): * embl (EST), Vers. 60 (16-SEP-1999)
  * * emblnew (EST), Vers. 60+ (18-OCT-1999) Using unmasked
  sequence
  * Minimum score: 60; Minimum identity: 70 %;
  > 'GSSs': BLASTN 2.0.9 (Altschul et al.)
  * Database(s): * embl (GSS), Vers. 60 (16-SEP-1999)
  * * emblnew (GSS), Vers. 60+ (18-OCT-1999) Using unmasked
  sequence
  * Minimum score: 60; Minimum identity: 70 %;
  > 'Tandem Repeats': GDE 2.2 option 'tandem'
  * Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
  * Treat N's as mismatches? YES; Allow uniform consensi? NO >
  > 'Inverted Repeats': GDE 2.2 option 'inverted'
  > 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
  Islands': GDE 2.2 option 'cpG'
  * CpG island region size 100 bp;
  * Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
  e-PCR (Schuler)
  * Margin: 50; Number of mismatches allowed: 0; Word size: 7
  STS database: 'dbSTS markers'
  > 'rRNA Scan': tRNAscan-SE (Lowie & Eddy), Vers. 1.11 Sequence
  overlaps with Acc.Nrs. AL050303 and AL078614.
  Location/Qualifiers
  1. 73465
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="21"
  /clone="RPCIP704M13256Q2"
  /map="21q11.1"
  (3.461)...(89.577)
  /note="match: ESTs AL042464 R98070 T90747 AL044865 T90457
  AA578081"
  complement((240.666)...(445.938))
  /note="match: ESTs AF000593 R98316 AA584792 AT1761237
  T83289 AJ003413"
  341...424
  /note="homology = 72.6%, counts = 6"
  /rpt_family="ggcctgatcagtg repeat"
  /rpt_type=TANDEM
  378...449
  /note="GENSCAN prediction, score = 6.32"
  394...420
  /note="XPOUND prediction, score = 0.213"
  649...728
  /note="GRAIL, score = 40%, comment = marginal"
  complement((730.891)...(923.1285))
  /note="match: GSSs AQ341870 AQ074436 AQ079640 AQ437414
  AQ475020"
  826...1098
  /note="match: EST AL042124"
  complement((6062.6213)...6315)
  /note="match: ESTs AT372769 AA915976 AJ003450"

```

```

repeat_region
  /note="IR1, 85% complementary to IR1' (1523..1795)"
  /rpt_type=INVERTED
  1523..1795
  /note="IR1', 85% complementary to IR1 (826..1098)"
  /rpt_type=INVERTED
  (1526.1986)..(1734.2126)
  /note="match: GSSs AQ440604 AQ074436 B85689 AQ079640
  AQ539584 AQ332219 AQ308154 AQ437414 AQ475020 AQ41870"
  (1683.1819)..(1795.2126)
  /note="match: ESTs AL044865 AA584792 AT1761237 AL040943
  AL043508"
  complement((1777.1781)..(1870.1902))
  /note="match: GSSs AQ610054 AQ428998"
  2128..(2234.2312)
  /note="match: ESTs AL043508 AL040943"
  2128..(2238.2327)
  /note="match: GSSs AQ308154 AQ539584 AQ332219"
  2137..2299
  /note="GENSCAN prediction, score = 8.01"
  2284..2299
  /note="XPOUND prediction, score = 0.356"
  2328..2560
  /note="match: EST AA349997"
  2328..(2408.2418)
  /note="match: GSSs AQ332219 AQ539584"
  complement(2462..2553)
  /note="MzEF prediction, score = 0.564"
  (2587.2703)..(3035.3037)
  /note="match: ESTs H23104 N45556"
  2687..2870
  /note="MzEF prediction, score = 0.709"
  complement((2697.3121)..(2939.3503))
  /note="match: ESTs AA927218 AI970509 AJ003449 AL040461
  AI216366 AA912892"
  2929..2944
  /note="XPOUND prediction, score = 0.605"
  3245..3320
  /note="homology = 76.3%, counts = 4"
  /rpt_family="attggggatccaggca repeat"
  /rpt_type=TANDEM
  (3438.3719)..(3801.3933)
  /note="match: ESTs AL042367 AL044856 AA316256 AL045430
  AA38573 AL042369"
  complement((3546.3822)..(3845.3870))
  /note="match: ESTs AW087494 AL040944 AI925550 AL040461
  AA961620 AA878363 AI806762 N48708 AI015474 AI826553"
  3653..3870
  /note="XPOUND prediction, score = 0.830"
  3682..3870
  /note="GRAIL, score = 71%, comment = good"
  4090..4179
  /note="match: EST 244176"
  complement((4316.4583)..(4561.4811))
  /note="match: ESTs AW026376 AA864338 AL040944 AI125870
  AI926636 AI126131 AI813897 AI925550 AL043509 AA970054
  AI457652"
  (4360.4364)..(4415.4500)
  /note="match: ESTs CI6613 AL044856 AL045430 AL042369
  AL042367 AI372770"
  complement(4367..4500)
  /note="match: GSS AQ188162"
  complement(4512..4661)
  /note="GRAIL, score = 71%, comment = good shadow"
  complement(4542..4693)
  /note="match: GSS AQ188162"
  (5664.5847)..(5859.5940)
  /note="match: ESTs AL046682 AL042369 AL042367 AA382548"
  complement((5684.5922)..(6164.6191))
  /note="match: GSSs AQ0809119 AQ322681 AQ771684"
  5937..6248
  /note="match: EST AL042124"
  complement((6062.6213)..6315)
  /note="match: ESTs AT372769 AA915976 AJ003450"

```

```

misc_feature (6210..6239) . (6248..6428)
/notes="match: GSS A0634548"
misc_feature 6759..6953
/notes="match: EST 244176"
exon 6925..7048
/notes="GENSCAN prediction, score = 0.54"
/notes="match: ESTs AL079705 R48206"
misc_feature (7011..7306) . (7157..7478)
/notes="match: ESTs AL079705 R48206"
misc_feature complement((7184..7229) . (7284..7298))
/notes="match: ESTs AI015474 AI145105 AI033019 AI222825
AI261513 AA915976 AJ003450 AI372769 AI674958 AA878363
AI806762 AI122718 N48708"
exon 7239..7298
/notes="XPOUND prediction, score = 0.503"
misc_feature complement(7383..7452)
/notes="match: EST AL079706"
misc_feature (7644..8094) . (7983..8186)
/notes="match: ESTs C00409 AA420700 AL040943 AL079705
AA420670"
misc_feature complement((7681..8303) . (7930..8604))
/notes="match: ESTs AW087637 AA860908 AA961620 AA846418
AI216366 AI674958 AA878363 AI806762 Z40157 AI122718 N48708
AA93884 AA912892 AA927218 AW026376 AW087494 AI863316
AI145105 AA420670 AA778653 AI033019 AI222825 AI261513"

Query Match 67.7%; Score 21; DB 92; Length 73465;
Best Local Similarity 82.8%; Pred. No. 89;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggacagataaaactgtggaata 30
||||| ||||| ||||| |||||
DB 43834 CAGCAGGTAAAGATAGATTGAAGATA 43862

RESULT 10
AL162851/c 152318 bp DNA HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 13 clone RP11-186M12, *** SEQUENCING IN
DEFINITION PROGRESS ***, 28 unordered pieces.
ACCESSION AL162851
VERSION AL162851.5 GI:9930860
KEYWORDS HTG; HTGS; PHASE1...
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Burton,J.
Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9926516.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project information
Center project name: bA186M12
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 132981 bases at least Q40
Consensus quality: 140365 bases at least Q30
Consensus quality: 145091 bases at least Q20
Insert size: 149618; sum-of-contigs
Quality coverage: 177713; 5.3% error; agarose-fp
Quality coverage: 2.24x in Q20 bases; sum-of-contigs Quality
coverage: 2.25x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2850: contig of 2850 bp in length
* 2851 2950: gap of 100 bp
* 2951 6292: contig of 3342 bp in length
* 6293 6392: gap of 100 bp
* 6393 9033: contig of 2641 bp in length
* 9034 9133: gap of 100 bp
* 9134 15105: contig of 5972 bp in length
* 15106 15205: gap of 100 bp
* 15206 21020: contig of 5815 bp in length
* 21021 21120: gap of 100 bp
* 21121 25318: contig of 4198 bp in length
* 25319 25418: gap of 100 bp
* 25419 28317: contig of 2899 bp in length
* 28318 28417: gap of 100 bp
* 28418 37932: contig of 9515 bp in length
* 37933 38032: gap of 100 bp
* 38033 40383: contig of 2351 bp in length
* 40384 40483: gap of 100 bp
* 40484 43573: contig of 3090 bp in length
* 43574 43673: gap of 100 bp
* 43674 52721: contig of 9048 bp in length
* 52722 52821: gap of 100 bp
* 52822 55850: contig of 3029 bp in length
* 55851 55950: gap of 100 bp
* 55951 64716: contig of 8766 bp in length
* 64717 64816: gap of 100 bp
* 64817 84621: contig of 19805 bp in length
* 84622 84721: gap of 100 bp
* 84722 87712: contig of 2991 bp in length
* 87713 87812: gap of 100 bp
* 87813 93264: contig of 5452 bp in length
* 93265 93364: gap of 100 bp
* 93365 100284: contig of 6920 bp in length
* 100285 100384: gap of 100 bp
* 100385 104611: contig of 4227 bp in length
* 104612 104711: gap of 100 bp
* 104712 107367: contig of 2656 bp in length
* 107368 107467: gap of 100 bp
* 107468 114106: contig of 6639 bp in length
* 114107 114206: gap of 100 bp
* 114207 118005: contig of 3799 bp in length
* 118006 118105: gap of 100 bp
* 118106 128798: contig of 10693 bp in length
* 128799 128898: gap of 100 bp
* 128899 131245: contig of 2347 bp in length
* 131246 131345: gap of 100 bp
* 131346 133727: contig of 2382 bp in length
* 133728 133827: gap of 100 bp
* 133828 135873: contig of 2046 bp in length
* 135874 135973: gap of 100 bp
* 135974 144877: contig of 8904 bp in length
* 144878 144977: gap of 100 bp
* 144978 148859: contig of 3882 bp in length
* 148860 148959: gap of 100 bp
* 148960 152318: contig of 3359 bp in length.

```

```

FEATURES
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="13"
        /clone_lib="RP11-186M12"
        /clone_lib="RP11-186M12"
        /clone_lib="RP11-186M12"
    misc_feature
        1..2850
        /note="assembly fragment:00886
        fragment_chain:1"
    misc_feature
        2951..6292

```

```

/note="assembly_fragment:00453
fragment_chain:1"
misc_feature
/note="assembly_fragment:00046
fragment_chain:1"
misc_feature
/note="assembly_fragment:00325
fragment_chain:1"
misc_feature
/note="assembly_fragment:00973
fragment_chain:1"
misc_feature
/note="assembly_fragment:00197
fragment_chain:2"
misc_feature
/note="assembly_fragment:00937
fragment_chain:2"
misc_feature
/note="assembly_fragment:00369
fragment_chain:2"
misc_feature
/note="assembly_fragment:00014
fragment_chain:3"
misc_feature
/note="assembly_fragment:00726
fragment_chain:3"
misc_feature
/note="assembly_fragment:00081
fragment_chain:4"
misc_feature
/note="assembly_fragment:00331
fragment_chain:4"
misc_feature
/note="assembly_fragment:00876
fragment_chain:5"
misc_feature
/note="assembly_fragment:00679
fragment_chain:5"
misc_feature
/note="assembly_fragment:00003"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00048"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00088"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00101"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00142"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00256"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00310"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00335"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00511"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00657"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00664"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00779"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00823"
fragment_chain:5"
misc_feature
/note="assembly_fragment:00373
fragment_chain:5"
vector_side:right"
clone_end:T7

```

BASE COUNT 42602 a 31886 c 31286 g 43830 t 2714 others
ORIGIN

Query Match

67.7%; Score 21; DB 79; Length 152318;

```

Best Local Similarity 82.8%; Pred. No. 91;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cagcaggagacagataaaactgtggaata 30
||||| | ||||| | | |||||
Db 28988 CAGCAGTAAGAATAAGATTATGGAATA 28960
||||| | ||||| | | |||||

RESULT 11
HS214G14/c
LOCUS HS214G14 168075 bp DNA HTG 08-FEB-2000
DEFINITION Homo sapiens chromosome 21 clone RCP1704G1421402 map 21q11.1, ***
SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AL050303
VERSION AL050303.2 GI:6981890
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 168075)
Nordsiek,G., Conrad,A., Dose,S., Grimm,M., Groet,J., Hornischer,K.,
Loehnert,T.H., Nizetic,D., Scharfe,M., Schoen,O., Yaspo,M.L. and
Bloeker,H.
Direct Submission
Submitted (25-MAY-1999) GBF, Dept. of Genome Analysis, Mascheroder
Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
nordsiek@gbf.de, bloecker@gbf.de
On Feb 16, 2000 this sequence version replaced gi:6966957.
All annotations in this database entry are developed by
computational tools. It is therefore not explicitly noted in the
feature lines that evidence is not experimental.
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'AnnoMitter' (Hornischer & Bloeker).
+ Programs used by 'AnnoMitter':
+ GeneFinder (Green), Vers. 084
+ Organism: human
+ GenScan (Burge & Karlin), Vers. 1.0
+ Used matrix: vertebrate; Minimum score: 0
+ Grail (Xu et al.), Vers. 1.3
+ Organism: human
+ Mzef (Zhang)
+ Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
+ & Skolnick)
+ Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
+ BLASTN 2.0.9 (Altschul et al.)
+ Database(s): * RepBase (human), released 22-DEC-1995
+ * RepBase (mammal), released 22-DEC-1995
+ Minimum identity: 70 %;
+ * ESTs: BLASTN 2.0.9 (Altschul et al.)
+ * Database(s): * embl (EST), Vers. 60 (16-SEP-1999)
+ * emblnew (EST), Vers. 60+ (15-DEC-1999)
+ Using sequence with
+ masked repeats
+ Minimum score: 60; Minimum identity: 90 %;
+ * 'GSSs': BLASTN 2.0.9 (Altschul et al.)
+ * Database(s): * embl (GSS), Vers. 60 (16-SEP-1999)
+ * emblnew (GSS), Vers. 60+ (15-DEC-1999)
+ Using sequence with
+ masked repeats
+ Minimum score: 60; Minimum identity: 90 %;
+ * 'Random Repeats': GDE 2.2 option 'tandem'
+ Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
+ Treat N's as mismatches? YES; Allow uniform consensi? NO >
+ 'Inverted Repeats': GDE 2.2 option 'inverted'
+ 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
+ Islands': GDE 2.2 option 'cpg'
+ CPG island region size 100 bp;
+ Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
+ e-PCR (Schuler)

```

```

Margin: 50; Number of mismatches allowed: 0; Word size: 7
SIS database: 'dbSTS markers'
> 'trna Scan': 'trnascan-se (Lowe & Eddy), Vers. 1.11.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 15941: contig of 15941 bp in length
* 15942 16741: gap of 800 bp
* 16742 168075: contig of 151334 bp in length.
FEATURES
    source
        1. 168075
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q11.1"
            /clone="RPCIP704G14214Q2"
        repeat_region
            1. 223
                /note="91% identity: matches 59. .280 of consensus"
                /rpt_family="AluSb"
        exon
            1. 41
                /note="XPOUND prediction, score = 0.421"
                /complement(1. .223)
                /note="87% identity: matches 245. .468 of consensus"
                /rpt_family="L1"
        repeat_region
            254. .313
                /note="match: GSS AQ118599"
                /complement(304.687) . (614.739)
        misc_feature
            304.687 . (614.739)
                /note="match: ESTs AI476532 N32072 AAL76985 H55081
                AI651642"
        misc_feature
            complement((448.694) . (699.739))
                /note="match: ESTs AI242520 AA927847 AA588848 AA884098
                AI041215 AI082168 AI936548 AA815249 AI911046 AA658461
                N45384 AI797236"
        exon
            576. .614
                /note="MZEUF prediction, score = 0.879"
        exon
            603. .614
                /note="XPOUND prediction, score = 0.277"
        misc_feature
            680. .727
                /note="match: GSS AQ371796"
                /complement(682. .729)
        misc_feature
            /note="match: GSS AQ418151"
                /complement((792.793) . (868.898))
        misc_feature
            /note="match: ESTs AI659600 AA927847 AA588848 AA884098
            AI041215 AI082168 AI936548 AA815249 AI911046 AA658461
            793. . (888.898)
        misc_feature
            /note="match: ESTs AI651642 AI200967 AA578971"
        repeat_region
            899. .929
                /note="IR1, 83% complementary to IR1' (943. .973)"
                /rpt_type=INVERTED
        repeat_region
            943. .973
                /note="IR1', 83% complementary to IR1 (899. .929)"
                /rpt_type=INVERTED
        misc_feature
            /complement((991.1044) . (1207.1223))
                /note="match: GSSs AQ392269 AQ311426 AQ310544 AQ007480
                AQ392185"
        misc_feature
            /note="match: EST AA937270"
                /complement(1234. .1328)
        repeat_region
            1042. .1125
                /note="91% identity: matches 4615. .4709 of consensus"
                /rpt_family="L1"
        repeat_region
            complement(1379. .1954)
                /note="88% identity: matches 431. .1006 of consensus"
                /rpt_family="L1"
        repeat_region
            complement(1972. .2023)
                /note="94% identity: matches 3197. .3249 of consensus"
                /rpt_family="L1"
        repeat_region
            complement(2076. .2292)
                /note="91% identity: matches 187. .403 of consensus"

```

```

/rpt_family="L1"
2201. .2293
/note="XPOUND prediction, score = 0.707"
2331. .2820
/note="89% identity: matches 2701. .3192 of consensus"
/rpt_family="L1"
2865. .2995
/note="IR5, 85% complementary to IR5' (4561. .4691)"
/rpt_type=INVERTED
2872. .2995
/note="IR4, 87% complementary to IR4' (4561. .4684)"
/rpt_type=INVERTED
2875. .2995
/note="IR2, 82% complementary to IR2' (3283. .3403)
IR3, 89% complementary to IR3' (4561. .4681)"
/rpt_type=INVERTED
complement((2930.2934) . (2983.2990))
/note="match: GSSs AQ471711 AQ041712 AQ410967 AQ034367
AQ219103 B55537 AQ185675 AQ019038"
(2933.2947) . (2979.3025)
/note="match: ESTs AW188087 AA577992 AI979227 AI623755
AA423972 AL048971 T35301 AI659905 AI927405 F01778 AI963294
AA883246 AA824522 AA932573 AI119402 AA708022 AI300484
AA931704 AI697170 AA448484 Z38954 AI079687 AI436442 F01552
AA226611 AA641060 AI191977 AA962425 AA989247 AI341212
AA657489 AI799492 AI247389 AA911085 AI979195 AA770179
AI867437 AI798417 AA262248 AA987270 AI458178 AI701161
AW193871"
complement(2933. .2987)
/note="match: ESTs AA262235 AA432049"
2986. .3024
/note="match: GSS AQ041712"
complement((3067.3141) . (3192)
/note="match: GSSs AQ207681 AQ185675 B30835"
3152. .3192
/note="match: GSS B18150"
(3246.3308) . (3300.3354)
/note="match: GSSs AQ471711 AQ027470 AQ672190 AQ007282
B30835"
(3283.3291) . (3345.3354)
/note="match: ESTs AA262235 AA318324 AA432049"
complement((3283.3299) . (3344.3351))
/note="match: ESTs AW188087 T02820 AI979227 AI623755
AA423972 AL048971 T35301 AI659905 AI927405 F01778 AI963294
AA883246 AA824522 AA932573 AI119402 AA708022 AI300484
AA931704 AI697170 AA448484 Z38954 AI079687 AI436442 F01552
AA226611 AA641060 AI191977 AA962425 AA989247 AI341212
AA657489 AI799492 AI247389 AA911085 AI979195 AA770179
AI867437 AI798417 AA262248 AA987270 AI458178 AI701161
AW193871"
2283. .3403
/note="IR2', 82% complementary to IR2 (2875. .2995)"
/rpt_type=INVERTED
3401. .3502
/note="match: EST T02820"
complement((3445.3569) . (3524.3623))
/note="match: GSSs AQ562833 AQ567848 AQ185675 B30835
AQ041712 AQ410967"
(3571.3572) . (3612.3625)
/note="match: ESTs AW188087 AI979227 AI623755 AA423972
AL048971 T35301 AI659905 AI927405 F01778 AI963294 AA883246
AA824522 AA932573 AI119402 AA708022 AI300484 AA931704
AI697170 AA448484 Z38954 AI079687 AI436442 F01552 AA226611
AA641060 AI191977 AA962425 AA989247 AI341212 AA657489
AW193871"
67.7% Score 21; DB 84; Length 168075;
Query Match
Best local similarity 82.8%; Pred. No. 92;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 cagcagcagacataaaactgtggaata 30
|||||
Db 163286 CAGCAGGTAAAGATAGAAATTATGGAATA 163258

```

RESULT 12

HS21C004/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

source

repeat_region

HS21C004 340000 bp DNA PRI 24-MAY-2000
 Homo sapiens chromosome 21 segment HS21C004.
 AL163204 AP001659 BA000005
 AL163204.2 GI:7717247

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 340000)
 Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
 Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
 Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
 Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
 Asakawa,S., Shintani,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
 Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
 Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
 Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
 Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
 Gardiner,K., Nizetic,D., Francis,F., Leirach,H., Reinhardt,R. and
 Yaspo,M.L.

Direct Submission
 Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
 Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
 Group * Institute of Molecular Biotechnology, Genome Analysis *
 Keio University School of Medicine, Dept. of Molecular Biology *
 GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
 Genetics (addresses see below)
 The Chromosome 21 Mapping and Sequencing Consortium consists of
 * RIKEN Genomic Sciences Center, Human Genome Research Group, *
 Sagamihara 228-8555, Japan.
 * e-mail: sakaki@gs.riken.go.jp
 * URL: http://hgp.gs.riken.go.jp/
 and
 * Institute of Molecular Biotechnology, Genome Analysis, *
 Beutenbergstrasse 11, D-07745 Jena, Germany,
 * e-mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/
 and
 * Keio University School of Medicine, Dept. of Molecular Biology, *
 Tokyo 160-8582, Japan.
 * e-mail: shimizu@mb-med.keio.ac.jp
 * URL: http://adenine.dmb.med.keio.ac.jp/
 and
 * GBF, Dept. of Genome Analysis,
 * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e-mail:
 info.genome@gbf.de
 * URL: http://genome.gbf.de/
 and
 * Max-Planck Institute for Molecular Genetics,
 * Ihnestrasse 73, D-14195 Berlin, Germany,
 * e-mail: info-chr21@molgen.mpg.de
 * URL: http://chr21.rz-berlin.mpg.de/
 Location/Qualifiers
 1..340000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q11.1"
 <1..3300
 /note="Accession No. AP001465"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q11.1"
 /clone="R663H15, 5' partial"
 /clone_lib="RPC1-11 BAC library"
 473..1061

/note="L1PAL2"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 complement(1115..1235)
 /note="L1PAL3"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 1228..2242
 /note="L1PAL2"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 2243..2552
 /note="Alu"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 2301..170825
 /note="Accession No. AL050303"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q11.1"
 /clone_lib="RPC11,3-5 PAC library"
 2553..2567
 /note="L1PAL2"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 2569..3516
 /note="MER4-internal"
 /rpt_family="LTR/MER4-group"
 /rpt_type=DISPERSED
 complement(3535..4595)
 /note="L1PA7"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 4622..5134
 /note="L1P"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 complement(8986..9422)
 /note="L1P"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 complement(9593..9894)
 /note="AluJo"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 10899..11367
 /note="L1M1"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 11506..11811
 /note="AluSg"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 complement(11883..12556)
 /note="L1MEC"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 complement(12902..13457)
 /note="L1MCC"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 13458..13487
 /note="(7C)n"
 /rpt_family="Simple_repeat"
 /rpt_type=TANDEM
 complement(13488..13708)
 /note="L1MCC"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 complement(14935..15084)
 /note="MIR"

Thu Oct 25 13:08:32 2001

us-09-462-955-1_copy_941_971.rge

```

repeat_region      /rpt_family="SINE/MIR"
                   /rpt_type=DISPERSED
                   15303..15409
                   /note="(TTCC)n"
repeat_region      /rpt_family="Simple_repeat"
                   /rpt_type=TANDEM
                   complement(15413..15735)
                   /note="AluJb"
repeat_region      /rpt_family="SINE/Alu"
                   /rpt_type=DISPERSED
                   complement(16256..16379)
                   /note="MLT1A1"
repeat_region      /rpt_family="LTR/MaLR"
                   /rpt_type=DISPERSED
                   complement(16626..16919)
                   /note="AluSp"
repeat_region      /rpt_family="SINE/Alu"
                   /rpt_type=DISPERSED
                   complement(17158..17447)
                   /note="AluSq"
repeat_region      /rpt_family="SINE/Alu"
                   /rpt_type=DISPERSED
                   complement(17603..17794)
                   /note="MIR"
repeat_region      /rpt_family="SINE/MIR"
                   /rpt_type=DISPERSED
                   18054..18154
                   /note="MLT2H"
repeat_region      /rpt_family="LTR/Retroviral"
                   /rpt_type=DISPERSED
                   complement(18708..19019)
                   /note="AluX"
repeat_region      /rpt_family="SINE/Alu"
                   /rpt_type=DISPERSED
                   19375..19481
                   /note="AluSg1"
repeat_region      /rpt_family="SINE/Alu"
                   /rpt_type=DISPERSED
                   complement(19492..19968)
                   /note="LMB3"
repeat_region      /rpt_family="LINE/L1"
                   /rpt_type=DISPERSED
                   20554..21037
                   /note="LIM4"
repeat_region      /rpt_family="LINE/L1"
                   /rpt_type=DISPERSED
                   21497..21749
                   /note="LTR37B"
repeat_region      /rpt_family="LTR/MBR4-group"
                   /rpt_type=DISPERSED
                   23106..23276
                   /note="(CA)n"
repeat_region      /rpt_family="Simple_repeat"
                   /rpt_type=TANDEM
                   23296..23470
                   /note="(CA)n"
repeat_region      /rpt_family="Simple_repeat"

```

```

Query Match      67.7%; Score 21; DB 92; Length 340000;
Best Local Similarity 82.8%; Pred.No. 94;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

OY 2 cagcaggacagaataaaactgtggaata 30
||||| ||||| ||||| ||||| |||||
Db 166036 CAGCAGGTAAGAAATAGAAATATGGAATA 166008

```

```

RESULT 13
AC021220      AC021220 214348 bp DNA      HTC      15-NOV-2000
LOCUS      Homo sapiens chromosome 4 clone RP11-530I17, WORKING DRAFT
DEFINITION      SEQUENCE, 38 unordered pieces.
ACCESSION      AC021220

```

VERSION KEYWORDS SOURCE ORGANISM

AC021220.4 GI:11178158
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 214348)
Waterston,R.H.
The sequence of Homo sapiens clone

Unpublished
2 (bases 1 to 214348)
Waterston,R.H.

Direct Submission
Submitted (15-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

On Nov 15, 2000 this sequence version replaced gi:9309528.

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0530I17
----- Summary Statistics -----
Sequencing vector: M13; 87%
Sequencing vector: plasmid; 13%
Chemistry: Dye-primer ET; 87% of reads
Chemistry: Dye-terminator Big Dye; 13% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195065 bases at least Q40
Consensus quality: 200600 bases at least Q30
Consensus quality: 203756 bases at least Q20
Insert size: 161000; agarose-fp
Insert size: 210648; sum-of-contigs
Quality coverage: 4.96 in Q20 bases; agarose-fp
Quality coverage: 4.62 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1584: contig of 1584 bp in length
* 1585 1684: gap of unknown length
* 1685 4194: contig of 2510 bp in length
* 4195 4294: gap of unknown length
* 4295 6424: contig of 2130 bp in length
* 6425 6524: gap of unknown length
* 6525 9844: contig of 3320 bp in length
* 9845 9945: gap of unknown length
* 9945 14101: contig of 4157 bp in length
* 14102 14201: gap of unknown length
* 14202 21279: contig of 7078 bp in length
* 21280 21346: gap of unknown length
* 21347 31347: gap of unknown length
* 31347 40526: contig of 9080 bp in length
* 40527 40626: gap of unknown length
* 40627 52862: contig of 12236 bp in length
* 52863 52963: gap of unknown length
* 52963 66090: contig of 13128 bp in length
* 66091 82659: contig of 16469 bp in length
* 82660 82759: gap of unknown length
* 82760 104797: contig of 22038 bp in length
* 104798 104897: gap of unknown length
* 104898 129319: contig of 24422 bp in length
* 129320 129419: gap of unknown length
* 129420 176511: contig of 47092 bp in length

```


CC	hypothetical: No;				
CC	anti-sense: No;				
FH	Key	Location/Qualifiers			
FH					
FT	source	1. .4621			
FT		/organism='Mus musculus domesticus', FT			
FT		/tissue_type='Testis',			
FT	CDS	1< .4621			
FT		/gene='Mea-2'.			
		Location/Qualifiers			
		1. .4621			
		/organism='Mus musculus'			
		/db_xref="taxon:10090"			
		1384 a 1058 c 1306 g 872 t			1 others
	BASE COUNT				
	ORIGIN				

```

Query Match          66.5%; Score 20.6; DB 10; Length 4621;
Best Local Similarity 85.2%; Score No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2   cagcagggcagacgaataaaaactgtggaa 28
          | | | | | | | | | | | | | | | |
Db      2739 CTGCAGGAGAGAGAAAGACTGTGGAA 2765

Search completed: October 24, 2001, 11:42:34
Job time: 6339 sec

```

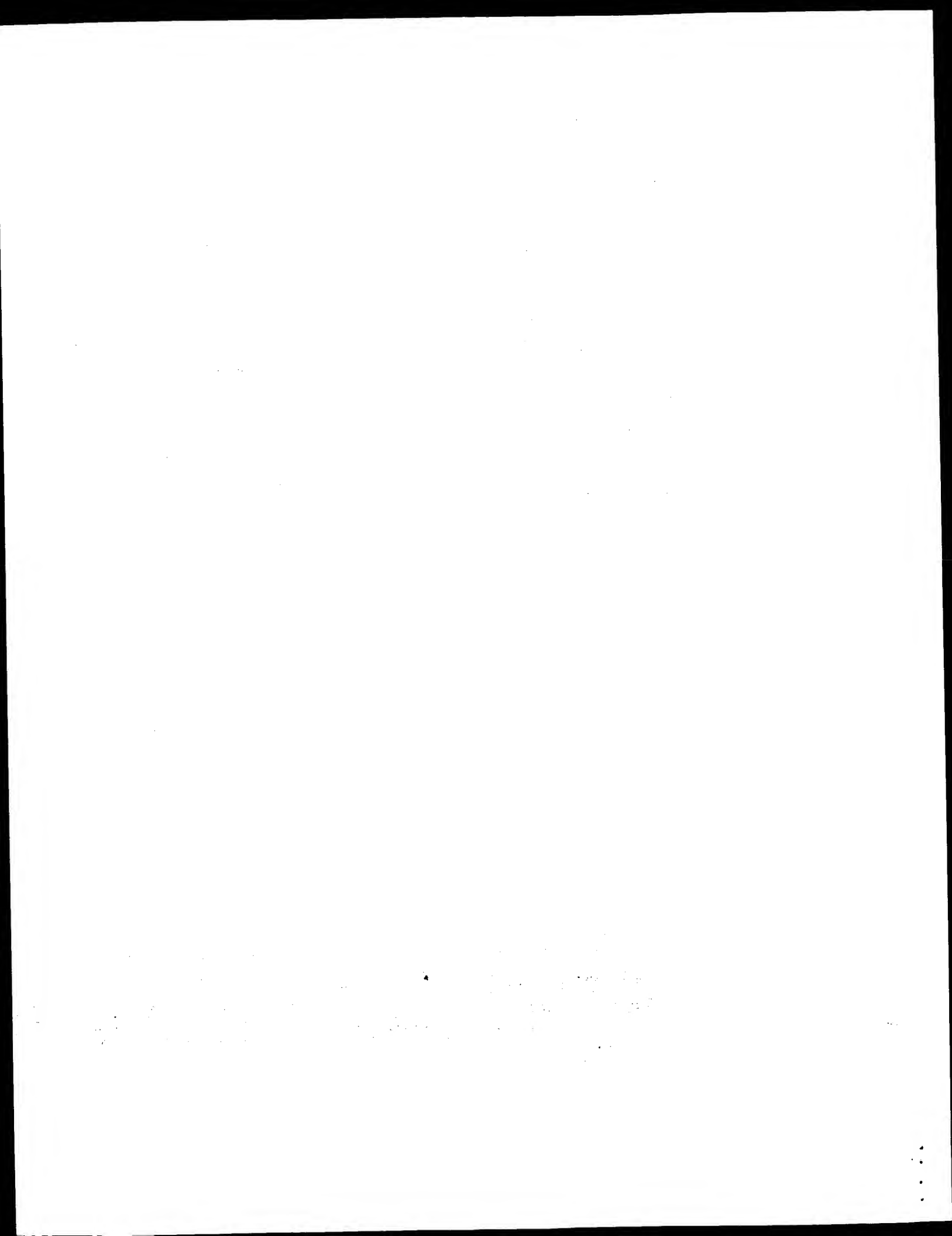
2 cagcaggacagaataaaaactgtggaa 28
| | | | | | | | | | | | | |
CTCCAGCCACACACAAGAACACTGTGGAA 216

RESULT	15	EL13044	4621 bp	DNA	PAT	24-JUN-1998
LOCUS		CDNA of mouse Mea-2 gens.				
DEFINITION		EL13044				
ACCESSION		EL13044.1 GI:3251856				
VERSION		JP 1997121869-A/1.				
KEYWORDS		Mus musculus.				
SOURCE		Mus musculus				
ORGANISM		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
REFERENCE		1 (bases 1 to 4621)				
AUTHORS		Kondo, M. and Sudou, S. .				
JOURNAL		MEA-2 GENE				
TITLE		Patent: JP 1997121869-A 1 13-MAY-1997;				
COMMENT		ITO HAM KK				
		OS Mus musculus domesticus (mouse)				
		PN JP 1997121869-A/1				
		PD 13-MAY-1997				
		PF 07-NOV-1995 JP 1995311638				
		PI KONDO MASAKI, SUDOU SHIZUO				
		PC C12N15/09;C07H21/02;C07H21/04;C07K14/47;C07K16/18;C12N1/21;C12Q1/68;				
		PC G01N33/53;G01N33/531;C12N1/21;C12R1:19;				
		CC G01N33/53;G01N33/531;C12N1/21;C12R1:19;				
		strandedness: Double;				
		CC topology: Linear;				

..
Thu Oct 25 13:08:32 2001

us-09-462-955-1_copy_941_971.rge

Page 19



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:06:56 ; Search time 393.52 Seconds
(without alignments)
49.464 Million cell updates/sec

Title: US-09-462-955-1-copy_941_971

Perfect score: 31
Sequence: 1 tcagcaggacagaataaaactgtggaatat 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.6	66.5	4621	AA74034	Mouse male-enhance
2	20.6	66.5	4621	AA74034	Mouse male enhance
3	19.8	63.9	232	AA04132	Human IRG27 relate
4	19.4	62.6	20844	AA081460	N. meningitidis pa
5	19.4	62.6	349980	AA721611	Neisseria meningit
6	19.4	62.6	837096	AA081489	N. meningitidis pa
7	19.2	61.9	802	AA044457	Human secreted exp
8	19.1	61.3	360	AA065352	Novel human polynu
9	19.1	61.3	404	AA065328	Novel human polynu
10	19.1	61.3	5578	AA076023	TGF-beta-2 promote
11	18.8	60.6	476	AA254436	Neisseria meningit

C 12	18.8	60.6	526	21	AAZ54437	Neisseria meningitidis
C 13	18.8	60.6	1001	21	AAAC57445	Arachidonic acid metabolism
C 14	18.8	60.6	1325	21	AAAC5143	Arabidopsis thaliana
C 15	18.8	60.6	2061	20	AA000614	Human secreted protein
C 16	18.8	60.6	2373	17	AAAT26997	N. meningitidis serotype 4
C 17	18.8	60.6	2376	17	AAAT26998	N. meningitidis serotype 4
C 18	18.8	60.6	2379	17	AAAT26999	N. meningitidis serotype 4
C 19	18.8	60.6	46593	21	AAAB1456	N. meningitidis serotype 4
C 20	18.8	60.6	349980	21	AAAF21612	Neisseria meningitidis
C 21	18.8	60.6	1830121	17	AAAT42063	Haemophilus influenzae
C 22	18.6	60.0	288	16	AAAT20038	Human gene signature
C 23	18.6	60.0	2095	21	AAAT55022	Membrane-bound protein
C 24	18.6	60.0	2095	22	AAAF92075	Human PRO1074 cDNA
C 25	18.6	60.0	2095	22	AAAF4168	Human PRO1074 cDNA
C 26	18.6	60.0	2189	20	AA087192	Human Dendritic cell
C 27	18.6	60.0	9253	22	AAAC82926	Human Dendritic cell
C 28	18.4	59.4	230	21	AAAC01770	S. enterica serovar
C 29	18.4	59.4	33778	22	AAAF38532	Human secreted protein
C 30	18.2	58.7	407	20	AAAF87632	Genomic fragment #
C 31	18.2	58.7	1095	21	AAAF37524	EST clone DY178.
C 32	18.2	58.7	1791	21	AAAF34297	Arabidopsis thaliana
C 33	18.2	58.7	1940	17	AAAT38690	Arabidopsis thaliana
C 34	18.2	58.7	3178	20	AAAX99865	Mouse CRTAM. Mus
C 35	18.2	58.7	16550	19	AAV21211	Human secreted protein
C 36	18.2	58.7	1664976	19	AAV21209	Methanococcus jannaschii
C 37	18.1	87350	18	AAAX83003	Methanococcus jannaschii	
C 38	17.8	57.4	627	21	AAAC43348	Human WRN genomic
C 39	17.8	57.4	854	21	AAAC48349	Arabidopsis thaliana
C 40	17.8	57.4	855	21	AAAC39482	Arabidopsis thaliana
C 41	17.8	57.4	861	21	AAAC33039	Arabidopsis thaliana
C 42	17.8	57.4	3361	20	AAAX80993	Arabidopsis thaliana
C 43	17.8	57.4	3813	22	AAAC91329	Mouse steroid receptor
C 44	17.6	56.8	294	21	AA051367	Human polynucleotide
C 45	17.6	56.8	702	21	AAAZ28695	Arabidopsis thaliana
						C. pneumoniae CPN1

ALIGNMENTS

RESULT 1
AAT74034
ID AAT74034 standard; cDNA: 4621 BP.
XX AC AAT74034;
XX AC AAT74034;
DT 16-SEP-1997 (first entry)
XX DE Mouse male-enhanced antigen-2 encoding cDNA.
XX DE Mouse; MEA-2; detecting mutation; ds.
KW Mouse; MEA-2; detecting mutation; ds.
XX OS Mus musculus domesticus.
XX FH Key Location/Qualifiers
CDS 570..4547
FT /*tag= a
FT /*product= Male-enhanced_antigen-2
FT misc_difference 307
FT /*tag= b
FT /*note= "Unspecified"
XX JP09121869-A.
XX PD 13-MAY-1997.
XX PF 07-NOV-1995; 95JP-0311638.
XX PR 07-NOV-1995; 95JP-0311638.
XX PA (ITOH-) ITO HAM KK.
XX DR WPI; 1997-314229/29.
XX DR P-PSDB; AAW19540.

Qy	1	tcagcaggacagataaaactgtggaatat	31
Db	69	tcagcagataaagagataaaactgtggaacat	99
RESULT	4		

AAA81460
 ID AAA81460 standard; DNA: 20844 BP.
 XX
 AC AAA81460;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_8 SEQ ID NO:8.
 XX
 DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX
 XX WO200022430-A2.
 PN
 XX
 XX 20-APR-2000.
 XX
 XX 08-OCT-1999; 99WO-US23573.
 XX
 XX 09-OCT-1998; 98US-0103794.
 PR
 XX 30-APR-1999; 99US-0132068.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 XX WPI: 2000-318079/27.
 DR
 XX
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX
 XX Claim 7: Page 320-326; 1760pp; English.
 XX
 XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 XX Sequence 20844 BP; 5244 A; 5005 C; 5645 G; 4949 T; 1 other;
 SQ

Query Match 62.6%; Score 19.4; DB 21; Length 20844;
 Best Local Similarity 79.3%; Pred. No. 65;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 cagcaggacagataaactgtggaata 30
 ||||| ||||| ||| | || || || |

Db 7682 cagcaggacagataaactgtggaata 7710
 RESULT 5
 AAF21611/c
 ID AAF21611 standard; DNA: 349980 BP.
 XX
 AC AAF21611;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.
 XX
 DE Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX
 XX WO2000066791-A1.
 PN
 XX
 XX 09-NOV-2000.
 PD
 XX
 XX 08-MAR-2000; 2000WO-US05928.
 PF
 XX
 XX 30-APR-1999; 99US-0132068.
 PR
 XX 08-OCT-1999; 99WO-US23573.
 PR
 XX 28-FEB-2000; 2000GB-0004695.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX
 XX (GENO-) INST GENOMIC RES.
 XX
 XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 XX
 XX WPI: 2000-647603/62.
 DR
 XX
 XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections -
 PT
 XX Claim 7: Appendix A; 692pp; English.
 CC
 CC The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.
 XX
 XX Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 other;
 SQ

Query Match 62.6%; Score 19.4; DB 21; Length 349980;
 Best Local Similarity 79.3%; Pred. No. 1e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Query Match 62.6%; Score 19.4; DB 21; Length 837096;
Best Local Similarity 79.3%; Pred. No. 1.1e-02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 cagcaggacagacataaaactgtggaata 30
|||||
Db 306539 CAGCAGGACATAATTACACGGTTGAAGA 306511

RESULT 7

AAA44457
ID AAA44457 standard; cDNA; 802 BP.

XX AAA44457;

AC AAA44457;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1032.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnerable; antiparkinsonian;
XX antitumor; osteoprotective; neuroprotective; nontropic; antiporiatic;
XX vaccine; anticonvulsant; antidepressant; gene therapy;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

OS WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.

XX 15-OCT-1998; 98US-0104436.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

XX Merberg D, Treacy M, Bowman MR;

XX WPI; 2000-317938/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (SESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -

PS Claim 1; Page 467; 803pp; English.

XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
XX sequence tags (SESTs), isolated from human, mouse, chicken and rat
XX tissue sources. The SESTs can have a range of activities depending on
XX the tissues they were isolated from. The activities include:
XX chemotactic; proliferative; immunomodulatory; haematopoietic;
XX chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
XX cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
XX antiasthmatic; vulnerable; antitumor; osteoprotective;
XX nontropic; antiparkinsonian; antiporiatic; cerebroprotective;
XX anticonvulsant; and antidepressant. The SESTs can be used for gene
XX therapy and in vaccines. The SESTs are useful as probes for the
XX identification and isolation of full-length cDNAs and genomic DNA
XX molecules which correspond to the SESTs. Proteins encoded by the SESTs
XX are useful in assays for determining biological activity and raising
XX antibodies. They may be useful for treatment of autoimmune disorders
XX (multiple sclerosis, insulin dependent diabetes), allergic conditions

QY 2 cagcaggacagacataaaactgtggaata 30
|||||
Db 241768 CAGCAGGACATAATTACACGGTTGAAGA 241740

RESULT 6

AAA81489/C
ID AAA81489 standard; DNA; 837096 BP.

XX AAA81489;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

OS WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHTR) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Scarlato V;
XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be
XX used in the diagnosis and treatment of N. meningitidis infection and
XX other Neisserial infections, for example, N.gonorrhoea -
XX Claim 7; Page 629-865; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies
XX against them, can be used in the manufacture (or in the manufacture of a
XX composition) for treating, preventing or diagnosing infection due to
XX Neisserial bacteria. For example, some of the identified proteins could
XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neissariae. Identification of sequences
XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious
XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX Multivalent vaccines have also been tried but none have successfully
XX overcome antigenic variability. The provision of further, complete
XX sequences may provide an opportunity to identify secreted or surface
XX exposed proteins that may be presumed targets for the immune system and
XX which are not antigenically variable or at least more conserved than
XX other more variable regions.

XX Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;

XX SQ

CC (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.

XX Sequence 802 BP; 270 A; 145 C; 170 G; 216 T; 1 other;

Query Match 61.9%; Score 19.2; DB 21; Length 802;
 Best Local Similarity 87.5%; Pred. No. 44;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 8 ggacagataaaactgtggaatat 31
 || ||||| ||||| ||||| |||||
 Db 668 gggcagataaaacagtgcaatat 691

RESULT 8
 AAF65352
 ID AAF65352 standard; cDNA; 360 BP.

XX AAF65352;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1108.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -
 XX Claim 9; Page 703; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to

CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

XX Sequence 360 BP; 123 A; 64 C; 72 G; 101 T; 0 other;

Query Match 61.3%; Score 19; DB 22; Length 360;
 Best Local Similarity 81.5%; Pred. No. 47;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cagcaggagacagataaaactgtggaa 28

|| ||||| ||||| ||||| |||||

Db 20 cagaaggatagataaaatgtaaaa 46

RESULT 9

AAF65328

ID AAF65328 standard; cDNA; 404 BP.

XX AAF65328;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1084.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -
 XX Claim 9; Page 699-700; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

XX Sequence 404 BP; 135 A; 72 C; 82 G; 112 T; 3 other;

Query Match 61.3%; Score 19; DB 16; Length 5578;
 Best Local Similarity 81.5%; Pred. No. 76;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 cagcgacagacgaataaactgtggaat 31
 ||||| ||||| ||||| ||||| |||||
 DB 4316 cagggcgagagtttaactgaggatct 4342

RESULT 11
 AAZ54436/c
 ID AAZ54436 standard; DNA: 476 BP.
 XX
 AC AAZ54436;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 923 partial DNA sequence SEQ ID NO:2821.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR P-PSDB; AAY75674.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 7; Page 1329; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 476 BP; 88 A; 125 C; 118 G; 145 T; 0 other;

Query Match 61.3%; Score 19; DB 22; Length 404;
 Best Local Similarity 81.5%; Pred. No. 48;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggacagacgaataaactgtgga 28
 ||||| ||||| ||||| ||||| |||||
 DB 20 cagaaggatagataaataatgtaaaa 46

RESULT 10
 AAQ76023
 ID AAQ76023 standard; cDNA: 5578 BP.
 XX
 AC AAQ76023;
 XX
 DT 10-AUG-1995 (first entry)
 XX
 DE TGF-beta-2 promoter region.
 XX
 KW TGF-beta-2; transforming growth factor-beta-2; promoter;
 KW raloxifene responsive element; osteoporosis; therapy; serum lipid;
 KW hypolipemic; antiestrogen; ss.
 XX
 OS Homo sapiens.
 XX
 PN
 XX
 FH Key
 FT misc_RNA
 FT
 FT Location/Qualifiers
 FT 1
 FT /*tag= a
 FT /note= "base at position 1 corresponds to
 FT TGF-beta-2 -227"
 FT
 FT TATA_signal
 FT 2248..2252
 FT /*tag= b
 FT 2278..3980
 FT /*tag= c
 FT 3981..5578
 FT /*tag= d
 FT 3635..3980
 FT /*tag= e
 FT /codon start= 1
 XX
 PN EP629697-A.
 XX
 PD 21-DEC-1994.
 XX
 PF 20-JUN-1994; 94EP-0304432.
 XX
 PR 21-JUN-1993; 93US-0081610.
 PR 18-MAY-1994; 94US-0246990.
 XX
 PA (ELIL) LILLY & CO ELI.
 XX
 PI Yang NN;
 XX
 PS WPI: 1995-024245/04.
 XX
 CC Identification of agents for the treatment of osteoporosis and
 CC serum lipid lowering agents - using nucleic acids from promoter
 CC regions of TGF-beta genes comprising regulatory elements
 CC designated raloxifene responsive elements.
 XX
 PS Disclosure; Page 29-32; 65pp; English.
 XX
 CC Promoter sequences from TGF-beta, -beta-2 and -beta-3, given in
 CC AAQ76022-24, respectively, were examined to identify regions
 CC responsive to the antiestrogen raloxifene. Plasmids encoding a
 CC reporter gene and partial sequences of the TGF-beta-3 promoter
 CC identified as raloxifene responsive elements were used to screen
 CC potential anti-osteoporosis agents.
 XX
 SQ Sequence 5578 BP; 1498 A; 1324 C; 1257 G; 1499 T; 0 other;

Best Local Similarity 76.7%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcagcagggacagataaaactgtggaata 30
||||| ||||| ||||| ||||| ||
Db 445 TCAGCGGGACAAAATAATATAGTGGATTA 416

RESULT 12
AAZ54437/c
ID AAZ54437 standard; DNA; 526 BP.
XX AC AAZ54437;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 923 partial DNA sequence SEQ ID NO:2823.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 09-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR P-PSDB; AAY75675.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 1330; 1453pp; English.

AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941 represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent PCR primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to Neisserial bacteria (e.g. meningitis and septicemia), to detect the presence of Neisseria bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.

Sequence 526 BP; 102 A; 135 C; 135 G; 154 T; 0 other;

Query Match 60.6%; Score 18.8; DB 21; Length 526;
Best Local Similarity 76.7%; Pred. No. 60;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 tcagcagggacagataaaactgtggaata 30
||||| ||||| ||||| ||||| ||
Db 495 TCAGCGGGACAAAATAATATAGTGGATTA 466

RESULT 13
AAC57445/c
ID AAC57445 standard; DNA; 1001 BP.
XX AC AAC57445;
XX
DT 25-JAN-2001 (first entry)
XX
DE Arachidonic acid metabolism related genomic biallelic marker #79.
XX
KW Human; biallelic marker; arachidonic acid metabolism; genotyping;
KW detection; hybridisation; phenotype; SNP; polymorphic base;
KW single nucleotide polymorphism; hybridisation assay; sequencing assay;
KW specific amplification assay; identification; ERBM; 12-LO-RBM;
KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
XX
OS Homo sapiens.
XX
PN WO200047771-A2.
XX
PD 17-AUG-2000.
XX
PF 11-FEB-2000; 2000WO-IB00184.
XX
PR 12-FEB-1999; 99US-0119917.
PR 23-MAR-1999; 99US-0275267.
PR 07-MAY-1999; 99US-0133200.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I;
PI WPI: 2000-571881/53.
XX
PT Novel biallelic markers useful for detecting conditions and genotypes
PT associated with arachidonic acid metabolism -
XX
PS Claim 13; Page 295; 802pp; English.

The present invention describes polynucleotides including biallelic markers derived from genes involved in arachidonic acid metabolism and from genomic regions flanking those genes. Methods from the present invention may be used to select individuals for clinical trials and predict responses to treatment with drugs. The polynucleotides may be used in hybridisation assays, sequencing assays and specific amplification assays for identifying an eicosanoid-related biallelic marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a segment of nucleotides containing an ERBM. The polynucleotides are useful in diagnostic kits. The markers may be used to detect conditions and genotypes associated with arachidonic acid metabolism. AAC57367 to AAC58018 and AAB24019 and AAB24020 represent sequences used in the exemplification of the present invention. N.B. Polymorphic bases (single nucleotide polymorphisms also known as SNPs) in the polynucleotide sequences from the present invention have been given as their corresponding degenerate bases e.g. a polymorphic base of C or T has been given as Y.

Sequence 1001 BP; 307 A; 202 C; 166 G; 325 T; 1 other;

Query Match 60.6%; Score 18.8; DB 21; Length 1001;
Best Local Similarity 76.7%; Pred. No. 68;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 cagcagggacagataaaactgtggaata 31
||||| ||||| ||||| ||||| ||

Thu Oct 25 13:08:33 2001

us-09-462-955-1_copy_941_971.rng

Db 596 CAGCAGTGACAGAGTGAACACAGTTTAATGT 567

RESULT 14
AAC45143
ID AAC45143 standard; DNA; 1325 BP.
XX AAC45143;
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 45452.
XX Arabidopsis thaliana.
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134321.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

```

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match Best Local Similarity 60.6%; Score 18.8; DB 21; Length 1325;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 2 cagcaggacagaaataaactgtggaatat 31
    | | | | | | | | | | | | | | | | | |
Db 1029 cggccggagctgaacaaactgtgttat 1058

```

RESULT 15
 AAX00614
 ID AAX00614 standard; DNA; 2061 BP.
 XX

```

AC AAX00614;
XX
DT 25-MAR-1999 (first entry)
XX
DE Human secreted protein gene 4 clone HAPBL78.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
XX WO9842738-A1.
XX
XX 01-OCT-1998.
XX
XX 19-MAR-1998; 98WO-0505311.
XX
XX 30-MAY-1997; 97US-0050937.
XX 21-MAR-1997; 97US-0041276.
XX 21-MAR-1997; 97US-0041277.
XX 21-MAR-1997; 97US-0041281.
XX 21-MAR-1997; 97US-0042344.
XX 30-MAY-1997; 97US-0048069.
XX 30-MAY-1997; 97US-0048094.
XX 30-MAY-1997; 97US-0048095.
XX 30-MAY-1997; 97US-0048096.
XX 30-MAY-1997; 97US-0048099.
XX 30-MAY-1997; 97US-0048131.
XX 30-MAY-1997; 97US-0048135.
XX 30-MAY-1997; 97US-0048154.
XX 30-MAY-1997; 97US-0048160.
XX 30-MAY-1997; 97US-0048186.
XX 30-MAY-1997; 97US-0048187.
XX 30-MAY-1997; 97US-0048188.
XX 30-MAY-1997; 97US-0048350.
XX 30-MAY-1997; 97US-0048351.
XX 30-MAY-1997; 97US-0048352.
XX 30-MAY-1997; 97US-0048355.
XX 05-AUG-1997; 97US-0054804.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Duan R, Ebner R, Ferrie AM, Florence KA;
XX Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
XX Rosen CA, Ruben SM, Shi Y, Young P;
XX
XX WPI: 1999-070066/06.
XX P-F5DB: AAW67810.
XX
XX New isolated human genes and the secreted polypeptides they encode -
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1: Page 170-171; 385pp; English.
XX
XX This sequence represents a nucleic acid molecule which encodes a
XX secreted human protein. The gene number, and the clone it is derived
XX from, are detailed in the descriptor line. The gene can be used to
XX generate fusion proteins by linking to the gene to a human immunoglobulin
XX Fc portion (e.g. AAX00602) for increasing the stability of the fused
XX protein as compared to the human protein only.
XX The invention relates to 87 novel genes and their fragments (nucleic
XX acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological
XX conditions can be diagnosed by determining the amount of the new
XX polypeptides in a sample or by determining the presence of mutations in

```

Thu Oct 25 13:08:33 2001

CC the new polynucleotides. Specific uses are described for each of the 87
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX00611 for described uses).

XX
 SQ Sequence 2061 BP; 665 A; 414 C; 494 G; 485 T; 3 other;

Query Match 60.6%; Score 18.8; DB 20; Length 2061;
 Best Local Similarity 76.7%; Pred. No. 77;
 Matches 23; Conservative: 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcagcagggaacagataaaactgttggaata 30
 |||| ||| |||| |||| |||| |||| ||||
 Db 428 tcagaaggacagaaaaatttgttggaaga 457

Search completed: October 24, 2001, 10:07:11
 Job time: 616 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:35 ; Search time 5479.82 Seconds
(without alignments)
53.476 Million cell updates/sec

Title: US-09-462-955-1_COPY_941_971

Perfect score: 31

Sequence: 1 tcagcaggacagacagataaaactgtggaatat 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_estl1:*

2: gb_est2:*

3: gb_est3:*

4: gb_est4:*

5: gb_est5:*

6: gb_est6:*

7: gb_est7:*

8: gb_est8:*

9: gb_est9:*

10: gb_est10:*

11: gb_est11:*

12: gb_est12:*

13: gb_est13:*

14: gb_est14:*

15: gb_est15:*

16: gb_est16:*

17: gb_est17:*

18: gb_est18:*

19: gb_est19:*

20: gb_est20:*

21: gb_est21:*

22: gb_est22:*

23: gb_est23:*

24: gb_est24:*

25: gb_est25:*

26: gb_est26:*

27: gb_est27:*

28: gb_est28:*

29: gb_est29:*

30: gb_est30:*

31: gb_est31:*

32: gb_est32:*

33: gb_est33:*

34: gb_est34:*

35: gb_est35:*

36: gb_est36:*

37: gb_est37:*

38: gb_est38:*

39: gb_est39:*

40: gb_est40:*

41: gb_est41:*

42: gb_est42:*

43: gb_est43:*

44: gb_est44:*

45: gb_est45:*

46: gb_est46:*

47: gb_est47:*

48: gb_est48:*

49: gb_est49:*

50: gb_est50:*

51: gb_est51:*

52: gb_est52:*

53: gb_est53:*

54: gb_est54:*

55: gb_est55:*

56: gb_est56:*

57: gb_est57:*

58: gb_est58:*

59: gb_est59:*

60: gb_est60:*

61: gb_est61:*

62: gb_est62:*

63: gb_est63:*

64: gb_est64:*

65: gb_est65:*

66: gb_est66:*

67: gb_est67:*

68: gb_est68:*

69: gb_est69:*

70: gb_est70:*

71: gb_est71:*

72: gb_est72:*

73: gb_est73:*

74: gb_est74:*

75: gb_est75:*

76: gb_est76:*

77: gb_est77:*

78: gb_est78:*

79: gb_est79:*

80: gb_est80:*

81: gb_est81:*

82: gb_est82:*

83: gb_est83:*

84: gb_est84:*

85: gb_est85:*

86: gb_est86:*

87: gb_est87:*

88: gb_est88:*

89: gb_est89:*

90: gb_est90:*

91: gb_est91:*

92: gb_est92:*

93: gb_est93:*

94: gb_est94:*

95: gb_est95:*

96: gb_est96:*

97: gb_est97:*

98: gb_est98:*

99: gb_est99:*

100: gb_est100:*

101: gb_est101:*

102: gb_est102:*

103: gb_est103:*

104: gb_est104:*

105: gb_est105:*

106: gb_est106:*

107: gb_est107:*

108: gb_est108:*

109: gb_est109:*

110: gb_est110:*

111: gb_est111:*

112: gb_est112:*

113: gb_est113:*

114: gb_est114:*

115: gb_est115:*

116: gb_est116:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCCGTCACGACG
Plate: 13 row: K column: 19
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
source
Location/Qualifiers
1..369
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

BASE COUNT 102 a 66 c 85 g 116 t
ORIGIN

Query Match 67.7%; Score 21; DB 136; Length 369;
Best Local Similarity 82.8%; Pred. No. 1.4e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggacagaataaaactgtggaata 30
||||| |||||| |||||| |||||
Db 150 CAGCAGGACAGCAATATCATCTGTCAATA 122

RESULT 4
BE483280/c
LOCUS 169387 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE483280
ACCESSION BE483280
VERSION BE483280.1 GI:9602813
KEYWORDS EST.
SOURCE COW.

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 510)
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
Mapping of Expressed Sequence Tags from a normalized bovine mammary
gland cDNA library
Unpublished (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@psi.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCCGTCACGACG
Plate: 13 row: L column: 19
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
source
Location/Qualifiers
1..510
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box B501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: polyT not found
High quality sequence stop: 316.

Location/Qualifiers
1..317
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:456352"
/clone_lib="Soares mouse placenta 4NBMP13.5 14.5"
/sex="unknown"
/tissue_type="placenta"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
15'.
TGTACCAATCTGAAGTGGAGCGCGCGCGGAAATTTTTTTTTTTTTTTTTT
T 3'; double stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 42 a 112 c 101 g 62 t
ORIGIN

Query Match 68.4%; Score 21.2; DB 20; Length 317;
Best Local Similarity 88.5%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 cagggacagaataaaactgtggaata 30
||||| |||||| |||||| |||||
Db 312 CAGGGACAGAACAAACGGTGGAAAA 287

RESULT 3
BE483265/c
LOCUS 169363 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE483265
ACCESSION BE483265
VERSION BE483265.1 GI:9602798
KEYWORDS EST.
SOURCE COW.

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 369)
Sonstegard,T.S., Capuco,A.V., Van Tassel,C.P., Ashwell,M.S. and
Wells,K.D.
Mapping of Expressed Sequence Tags from a normalized bovine mammary
gland cDNA library
Unpublished (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414

Library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."

BASE COUNT 131 a 110 c 106 g 163 t

Query Match .67.7%; Score 21; DB 136; Length 510;
Best Local Similarity .82.8%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cagcaggacagagaataaaactgtggaata 30
||||| ||||||| ||||| |||

Db 150 CAGCAGGACAGAGATATCACTGTCAATA 122

RESULT 5

AZ383240 587 bp DNA GSS 02-OCT-2000
LOCUS IM0140N13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0140N13 R, DNA sequence.

ACCESSION AZ383240
VERSION AZ383240.1 GI:10496940

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 587)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D. Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00
Plate: 0140 row: N column: 13

Seq primer: CACAGGAAACAGCTATGACC
Class: plasmid ends

High quality sequence stop: 587.
Location/Qualifiers
1..587

/organism="Mus musculus"
/strain="C57BL/6J"

/db_xref="taxon:10090"
/clone="UUGC1M0140N13"

/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and

FEATURES

source

purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 175 a 103 c 114 g 194 t

Query Match 67.7%; Score 21; DB 242; Length 587;
Best Local Similarity 82.8%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tcagcaggacagagaataaaactgtggaat 29
||||| ||||||| ||||||| |||

Db 504 TCAGAACTGACAGATAAACTGTGTATT 532

RESULT 6

AQ018224/c 222 bp DNA GSS 09-JUN-1998
LOCUS CIT-HSP-2306L9.TR CIT-HSP Homo sapiens genomic clone 2306L9, DNA
DEFINITION sequence.

ACCESSION AQ018224

VERSION AQ018224.1 GI:3196960

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 222)

AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

JOURNAL Unpublished (1998)

COMMENT Other GSSs: CIT-HSP-2306L9.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..222

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="2306L9"

/clone_lib="CIT-HSP"

/sex="Male"

/cell_type="Sperm"

/note="Vector: pBelOAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 67 a 52 c 34 g 69 t

Query Match

Best Local Similarity 67.1%; Score 20.8; DB 223; Length 222;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cagcaggacagagaataaaactgtg 25
|| ||||| ||||||| |||||

Db 219 CACCAGGAGAGATAAACTGTG 196

RESULT 7


```

DEFINITION RICS16437A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION D49255
VERSION D49255.1 GI:702964
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE
  1 (bases 1 to 416)
AUTHORS Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE Rice cDNA from callus 1995
JOURNAL Unpublished (1995)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp.
Location/Qualifiers
  1..416
  /organism="Oryza sativa"
  /strain="Nipponbare"
  /db_xref="taxon:4530"
  /clone_lib="Rice green shoot"
  /note="Green shoot (8 days old)"

BASE COUNT 83 a 111 c 139 g 82 t 1 others
ORIGIN

Query Match 66.5%; Score 20.6; DB 156; Length 416;
Best Local Similarity 85.2%; Pred. No. 2.6e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtgga 27
||||| ||||| ||||| ||||| |||||
Db 315 TCAGCAGCACAGGATAAAACTTTGGA 289

RESULT 10
AU096982/c
LOCUS AU096982 441 bp mRNA EST 30-JUN-2000
DEFINITION AU096982 Rice green shoot Oryza sativa cDNA clone si6790, mRNA
sequence.
ACCESSION AU096982
VERSION AU096982.1 GI:8859664
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

REFERENCE
  1 (bases 1 to 441)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from green shoot (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.
S16790_1A.
Location/Qualifiers
  1..441
  /organism="Oryza sativa"

```

```

/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="S16790"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"

BASE COUNT 90 a 120 c 147 g 84 t
ORIGIN

Query Match 66.5%; Score 20.6; DB 107; Length 441;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtgga 27
||||| ||||| ||||| ||||| |||||
Db 329 TCAGCAAGCACAGGATAAAACTTTGGA 303

RESULT 11
AU0144781/c
LOCUS AU0144781 441 bp DNA GSS 25-SEP-1998
DEFINITION HS_3092_A1_H01_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-3092 Col=1 Row=O, DNA sequence.
ACCESSION AU0144781
VERSION AU0144781.1 GI:3535434
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence tagged Connector
Plate: 3092 row: O column: 1
Class: BAC ends
High quality sequence stop: 441.
Location/Qualifiers
  1..441
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="Plate=3092 Col=1 Row=O"
  /clone_lib="CIT Approved Human Genomic Sperm Library D"
  /sex="male"
  /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 107 a 117 c 83 g 134 t
ORIGIN

Query Match 66.5%; Score 20.6; DB 224; Length 441;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 gcaggacagaataaaactgtggaata 30
||||| ||||| ||||| ||||| |||||
Db 78 GTAGTGTCAAGATAAAACTTTGGAATA 52

RESULT 12
D47654/c

```

Thu Oct 25 13:08:36 2001

```

LOCUS       D47654               462 bp    mRNA                      EST                02-AUG-1995
DEFINITION   R1C513277A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION    D47654
VERSION      D47654.1 GI:701363
KEYWORDS     EST.
SOURCE       Oryza sativa.
ORGANISM     Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1 (bases 1 to 462)
AUTHORS      Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE        Rice cDNA from callus 1995
JOURNAL      Unpublished (1995)
COMMENT      Contact: Takuji Sasaki
              National Institute of Agrobiological Resources
              Rice Genome Research Program
              2-1-2 Kannondai,Tsukuba
              Ibaraki,
              Japan 305
              Tel: 0298-38-7441
              Fax: 0298-38-7468
              Email: tsasaki@abr.affrc.go.jp.
              Location/Qualifiers
                1..462
                  /organism="Oryza sativa"
                  /strain="Nipponbare"
                  /db_xref="taxon:4530"
                  /clone_lib="Rice green shoot"
                  /note="green shoot (8 days old)"
BASE COUNT   92 a 123 c 151 g 95 t 1 others
ORIGIN
Query Match 66.5%; Score 20.6; DB 156; Length 462;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcagcagggacagaataaactgtgga 27
||||| ||||| ||||| ||||| |||||
Db 316 TCAGCAGGCACAGGATAAACTTTGGA 290

RESULT 13
LOCUS       AU089813/c          683 bp    mRNA                      EST                27-APR-2000
DEFINITION   AU089813 Rice green shoot Oryza sativa cDNA clone S16437, mRNA
sequence.
ACCESSION    AU089813
VERSION      AU089813.1 GI:7652248
KEYWORDS     EST.
SOURCE       Oryza sativa.
ORGANISM     Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1 (bases 1 to 683)
AUTHORS      Sasaki,T. and Yamamoto,K.
TITLE        Rice cDNA from green shoot (2000)
JOURNAL      Unpublished (2000)
COMMENT      Contact: Takuji Sasaki
              National Institute of Agrobiological Resources
              Rice Genome Research Program
              2-1-2 Kannondai,Tsukuba
              Ibaraki,
              Japan 305
              Tel: 0298-38-7441
              Fax: 0298-38-7468
              Email: tsasaki@abr.affrc.go.jp
              PROJECT = "RGP"
              S16437-87.
              Location/Qualifiers
                1..683
                  /organism="Oryza sativa"
                  /strain="Nipponbare"
                  /db_xref="taxon:4530"
                  /clone_lib="Rice green shoot"
                  /note="green shoot (8 days old)"
BASE COUNT   92 a 123 c 151 g 95 t 1 others
ORIGIN
Query Match 66.5%; Score 20.6; DB 156; Length 462;
Best Local Similarity 85.2%; Pred. No. 2.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcagcagggacagaataaactgtgga 27
||||| ||||| ||||| ||||| |||||
Db 316 TCAGCAGGCACAGGATAAACTTTGGA 290

FEATURES
source
  1..373
    /organism="Mentha x piperita"
    /cultivar="Black Mitcham"
    /db_xref="taxon:34256"
    /clone_lib="peppermint glandular trichome"
    /tissue_type="peltate glandular trichomes"
    /cell_type="secretory"
    /note="Vector: lambda ZAPII"
BASE COUNT   107 a 80 c 81 g 105 t
ORIGIN
Query Match 55.8%; Score 20.4; DB 113; Length 373;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 cagcagggacagaataaactgtggaat 31
||||| ||||| ||||| ||||| |||||
Db 237 CAGGACAGGATGGATCAACTGTGGGATAT 208

RESULT 15
LOCUS       BF515120          425 bp    mRNA                      EST                07-DEC-2000
DEFINITION   BF515120
              UI-H-BW1-anv-e-09-0-UI.s1 NC1_CGAP_Sub7 Homo sapiens cDNA clone
              IMAGE:3083632 3', mRNA sequence.
ACCESSION    BF515120
VERSION      BF515120.1 GI:11600299
KEYWORDS     EST.
  
```

```

/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone_lib="S16437"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"
BASE COUNT   188 a 152 c 143 g 196 t 4 others
ORIGIN
Query Match 66.5%; Score 20.6; DB 107; Length 683;
Best Local Similarity 85.2%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcagcagggacagaataaactgtgga 27
||||| ||||| ||||| ||||| |||||
Db 110 TCAGCAGGCACAGGATAAACTTTGGA 84

RESULT 14
LOCUS       AW255519/c          373 bp    mRNA                      EST                23-AUG-2000
DEFINITION   ML546 peppermint glandular trichome Mentha x piperita cDNA, mRNA
sequence.
ACCESSION    AW255519
VERSION      AW255519.1 GI:7244771
KEYWORDS     EST.
SOURCE       peppermint.
ORGANISM     Mentha x piperita
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Lamiales; Lamiales; Mentha.
REFERENCE    1 (bases 1 to 373)
AUTHORS      Lange,B.M., Wildung,M.R., Stauber,E.J., Sanchez,C., Pouchnik,D. and
              Croteau,R.
              Probing essential oil biosynthesis and secretion by functional
              evaluation of expressed sequence tags from mint glandular trichomes
              Proc. Natl. Acad. Sci. U.S.A. 97 (6), 2934-2939 (2000)
              20183992
              Contact: Lange, B.M.
              Institute of Biological Chemistry/Washington State University
              Pullman, WA
              Email: lange@mail.wsu.edu.
              Location/Qualifiers
                1..373
                  /organism="Mentha x piperita"
                  /cultivar="Black Mitcham"
                  /db_xref="taxon:34256"
                  /clone_lib="peppermint glandular trichome"
                  /tissue_type="peltate glandular trichomes"
                  /cell_type="secretory"
                  /note="Vector: lambda ZAPII"
BASE COUNT   107 a 80 c 81 g 105 t
ORIGIN
Query Match 55.8%; Score 20.4; DB 113; Length 373;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 cagcagggacagaataaactgtggaat 31
||||| ||||| ||||| ||||| |||||
Db 237 CAGGACAGGATGGATCAACTGTGGGATAT 208

RESULT 15
LOCUS       BF515120          425 bp    mRNA                      EST                07-DEC-2000
DEFINITION   BF515120
              UI-H-BW1-anv-e-09-0-UI.s1 NC1_CGAP_Sub7 Homo sapiens cDNA clone
              IMAGE:3083632 3', mRNA sequence.
ACCESSION    BF515120
VERSION      BF515120.1 GI:11600299
KEYWORDS     EST.
  
```

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 425)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library preparation: M.B. Soares Lab Clone Distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
source

Location/Qualifiers
 1..425
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3083632"
 /clone_lib="NCI_CGAP_Sub7"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7T3D-pac (Pharmacia) with a modified
 polylinker. Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub7
 is a subtracted library derived from NCI_CGAP_Sub6. The
 NCI_CGAP_Sub7 library had 12 million recombinants. A
 single-stranded DNA preparation of NCI_CGAP_Sub6 was used
 as a tracer in a subtractive hybridization with a driver
 comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
 1322376-1323911, 1456008-1456775, 1500552-1502855);
 NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
 (IMAGE Clones); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
 1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
 3851-3854 (IMAGE Clones); 1414920-1417991, 1520904-1522439
); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
 3733-3735 (IMAGE Clones); 1257096-1258631, 1459064-1470983
 , 1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
 2758-2759, 3062-3068 (IMAGE Clones); 985608-986759,
 1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
 LLAM 2644-2653, 2871-2872 (IMAGE Clones); 1057416-1061255
 , 1144584-1145351). (6% of the driver population), plus a
 pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE
 Clones); 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE
 Clones); 2710536-2712455) (4% of the driver population
), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE
 Clones); 2712456-2723591) (10% of the driver population),
 plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE
 Clones); 2723592-2729326) (40% of the driver population),
 plus a pool of 4032 clones from NCI_CGAP_Sub6 (IMAGE
 Clones); 2728969-2733190) (40% of the driver population).
 Subtraction was performed as previously described [Bonaldi
 , Lennon & Soares (1996): Normalization and Subtraction:
 Two Approaches to Facilitate Gene Discovery. Genome
 Research 6, 791-806.
 TAG_LIB=NCI_CGAP_Ov18
 TAG_TISSUE=ovary
 TAG_SEQ=GCACA"

BASE COUNT 127 a 91 g 137 t
 ORIGIN

Query Match 65.8%; Score 20.4; DB 149; Length 425;
 Best Local Similarity 80.0%; Pred. No. 3.2e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 tcacgagggacagataaaactgtggaata 30
 ||||| ||| | |||| | ||||| |||||

Db 61 TCAGTAGGCAAGAAAAAACTGTTGAATA 90

Search completed: October 24, 2001, 13:13:38
 Job time: 11803 sec

us-09-462-955-1_copy_941_971.rst

Thu Oct 25 13:08:36 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:12 ; Search time 180.6 Seconds
(without alignments)
32.495 Million cell updates/sec

Title: US-09-462-955-1-copy_941_971
Perfect score: 31
Sequence: 1 tcagcaggacagataaaactgtggaatat 31
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTGUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	61.3	5578	1	US-08-081-610-2
2	18.8	60.6	1957	4	US-09-352-990-11
3	18.8	60.6	2376	3	US-08-537-361E-3
4	18.8	60.6	2379	3	US-08-537-361E-5
5	18.8	60.6	3318	1	US-08-346-670A-1
6	18.8	60.6	3318	3	US-08-990-470A-1
7	18.8	60.6	3319	3	US-08-537-361E-1
8	18.2	58.7	1940	1	US-08-429-742-3
9	18.2	58.7	4550	4	US-09-103-663-35
10	18.2	58.1	87350	3	US-08-781-891-79
11	17.6	56.8	2653	1	US-08-325-553-1
12	17.6	56.8	2653	2	US-08-394-152A-1
13	17.6	56.8	3879	1	US-08-530-010-1
14	17.6	56.8	3879	1	US-08-484-101B-1
15	17.2	55.5	1094	4	US-09-173-914-5
16	17.2	55.5	1500	1	US-08-117-083-67
17	17.2	55.5	4195	1	US-08-340-011-1
18	17.2	55.5	4195	3	US-08-901-710-1
19	17.2	55.5	4416	3	US-08-795-430-1
20	17.2	55.5	4425	1	US-08-222-616-31
21	17.2	55.5	4425	5	PCT-US95-04228-31
22	17.2	55.5	4795	1	US-08-340-011-3
23	17.2	55.5	4795	3	US-08-901-710-3
24	17.2	55.5	9108	5	PCT-US95-04228-45
25	17	54.8	2073	4	US-09-032-523-6
26	17	54.8	2406	1	US-08-586-272-1
27	17	54.8	2406	3	US-09-082-969-1

28	17	54.8	2498	4	US-09-041-236-1	Sequence 1, Appli
c 29	17	54.8	13417	2	US-08-637-759B-37	Sequence 37, Appl
c 30	17	54.8	13417	3	US-08-871-355A-37	Sequence 37, Appl
c 31	17	54.8	51952	3	US-08-947-823-1	Sequence 1, Appli
c 32	16.8	54.2	366	2	US-08-672-345C-87	Sequence 87, Appl
c 33	16.8	54.2	366	2	US-08-672-345C-89	Sequence 89, Appl
c 34	16.8	54.2	720	2	US-07-956-399-1	Sequence 1, Appli
c 35	16.8	54.2	1716	3	US-08-656-034-9	Sequence 9, Appli
c 36	16.8	54.2	1816	2	US-08-951-148-2	Sequence 2, Appli
c 37	16.8	54.2	1816	2	US-09-165-234-2	Sequence 2, Appli
c 38	16.8	54.2	1816	3	US-09-274-570-2	Sequence 2, Appli
c 39	16.8	54.2	2160	3	US-08-656-034-1	Sequence 2, Appli
c 40	16.8	54.2	2562	1	US-08-146-421-4	Sequence 1, Appli
c 41	16.6	53.5	264	4	US-08-991-789A-24	Sequence 4, Appli
c 42	16.6	53.5	1240	2	US-08-874-186-39	Sequence 24, Appli
c 43	16.6	53.5	2240	2	US-08-671-978A-5	Sequence 39, Appli
c 44	16.6	53.5	2953	2	US-08-859-201-1	Sequence 5, Appli
c 45	16.6	53.5	3286	4	US-09-211-417-2	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-081-610-2
; Sequence 2, Application US/08081610
; Patent No. 5445941
; GENERAL INFORMATION:
; APPLICANT: Yang, Na N
; TITLE OF INVENTION: Materials and Methods for Screening
; TITLE OF INVENTION: Anti-Osteoporosis or Serum Lipid Lowering Agents
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti and Witcoff, Ltd.
; STREET: 10 S. Wacker Dr.
; CITY: Chicago
; STATE: IL
; COUNTRY: U.S.A
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/081.610
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Heaphy, Barbara A
; REGISTRATION NUMBER: 34,619
; REFERENCE/DOCKET NUMBER: 93,402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5578 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 2248..2252
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2278..3980
; FEATURE:
; NAME/KEY: intron
; LOCATION: 3981..5578
; FEATURE:
; NAME/KEY: misc_rna

us-09-462-955-1_copy_941_971.rni

Thu Oct 25 13:08:35 2001

LOCATION: 3635..3980
OTHER INFORMATION: /note= "CDS, Codon start = 1"
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 1..2
OTHER INFORMATION: /note= "Number 1 corresponds to
OTHER INFORMATION: TGFB-2 -2277"
US-08-081-610-2

Query Match 61.3%; Score 19; DB 1; Length 5578;
Best Local Similarity 81.5%; Pred. No. 14; 5; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 5 caggcagacagaataaaactgtggaatat 31
||||| ||||| ||||| ||||| |||||

Db 4316 CAGGGCAGAGTAACTGAGGAATCT 4342

RESULT 2

US-09-352-990-11
Sequence 11, Application US/09352990
Patent No. 6255090

GENERAL INFORMATION:
APPLICANT: Ramodu, Layo O.

APPLICANT: Orozco, Buddy

APPLICANT: Rafalski, Antoni

TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase

FILE REFERENCE: BB-1191

CURRENT APPLICATION NUMBER: US/09/352,990

CURRENT FILING DATE: 1999-07-14

EARLIER APPLICATION NUMBER: 60/092,866

EARLIER FILING DATE: July 15, 1998

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Microsoft Office 97

SEQ. ID. NO. 11

LENGTH: 1957

TYPE: DNA

ORGANISM: Oryza sativa

US-09-352-990-11

Query Match 60.6%; Score 18.8; DB 4; Length 1957;
Best Local Similarity 76.7%; Pred. No. 14;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 cagcaggacagaataaaactgtggaatat 31
||||| ||||| ||||| ||||| |||||

Db 1323 cagcaggacagaataaaactgtggaatat 1352

RESULT 3

US-08-537-361E-3

Sequence 3, Application US/08537361E

Patent No. 6121037

GENERAL INFORMATION:

APPLICANT: Stojiljkovic, Igor

APPLICANT: So, Magdalene

APPLICANT: Hwa, Vivian

APPLICANT: Heffron, Fred

APPLICANT: Nassif, Xavier

TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

STREET: 300 South Wacker Drive, 32nd Floor

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/537,361E

FILING DATE: 02-OCT-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: No. 6121037han, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 94,784-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-0001

TELEFAX: 312-913-0002

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2376 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2373

US-08-537-361E-3

Query Match 60.6%; Score 18.8; DB 3; Length 2376;
Best Local Similarity 76.7%; Pred. No. 15;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 tcagcaggacagaataaaactgtggaata 30
||||| ||||| ||||| ||||| |||||

Db 828 TCAGCAGGGCAGTAACTGAGGTTGAAGA 857

RESULT 4

US-08-537-361E-5

Sequence 5, Application US/08537361E

Patent No. 6121037

GENERAL INFORMATION:

APPLICANT: Stojiljkovic, Igor

APPLICANT: So, Magdalene

APPLICANT: Hwa, Vivian

APPLICANT: Heffron, Fred

APPLICANT: Nassif, Xavier

TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

STREET: 300 South Wacker Drive, 32nd Floor

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/537,361E

FILING DATE: 02-OCT-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: No. 6121037han, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 94,784-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-0001

TELEFAX: 312-913-0002

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 2379 base pairs
TYPE: nucleic acid.
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2376
US-08-537-361E-5

Query Match 60.6%; Score 18.8; DB 3; Length 2379;
Best Local Similarity 76.7%; Pred. No. 15;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtggaata 30
||||||| ||| ||| ||| ||| ||| |||
DB 828 TCAGCAGGGCATAATTACACGGTTGAAGA 857

RESULT 5
US-08-326-670A-1
; Sequence 1, Application US/08326670A
; Patent No. 5698438
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: A No. 5698438el Bacterial Hemoglobin Receptor Gene and Uses
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/326.670A
FILING DATE: 18 OCT 1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 5698438nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3318 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 470..2848
US-08-326-670A-1

Query Match 60.6%; Score 18.8; DB 1; Length 3318;
Best Local Similarity 76.7%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtggaata 30
||||||| ||| ||| ||| ||| ||| |||
DB 1297 TCAGCAGGGCATAATTACACGGTTGAAGA 1326

RESULT 6
US-08-990-470A-1
; Sequence 1, Application US/08990470A
; Patent No. 6123942
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6123942el Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehrnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/990.470A
FILING DATE: 15-DEC-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6123942nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-K
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3318 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 470..2845
US-08-990-470A-1

Query Match 60.6%; Score 18.8; DB 3; Length 3318;
Best Local Similarity 76.7%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcagcaggacagaataaaactgtggaata 30
||||||| ||| ||| ||| ||| ||| |||
DB 1297 TCAGCAGGGCATAATTACACGGTTGAAGA 1326

RESULT 7
US-08-537-361E-1
; Sequence 1, Application US/08537361E
; Patent No. 6121037
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian

```

APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 471..2848
US-08-537-361E-1

Query Match 60.68; Score 18.8; DB 3; Length 3319;
Best Local Similarity 76.78; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcacgaggacagaataaaactgtggaata 30
||||| ||||| ||||| |||||
Db 1298 TCACGAGGGCATAATTACAGGTTGAAGA 1327

RESULT 8
US-08-429-742-3
; Sequence 3, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; OPERATING SYSTEM: Version #1.25

APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 471..2848
US-08-537-361E-1

Query Match 60.68; Score 18.8; DB 3; Length 3319;
Best Local Similarity 76.78; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcacgaggacagaataaaactgtggaata 30
||||| ||||| ||||| |||||
Db 1298 TCACGAGGGCATAATTACAGGTTGAAGA 1327

RESULT 8
US-08-429-742-3
; Sequence 3, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; OPERATING SYSTEM: Version #1.25

APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 471..2848
US-08-537-361E-1

Query Match 60.68; Score 18.8; DB 3; Length 3319;
Best Local Similarity 76.78; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcacgaggacagaataaaactgtggaata 30
||||| ||||| ||||| |||||
Db 1298 TCACGAGGGCATAATTACAGGTTGAAGA 1327

RESULT 8
US-08-429-742-3
; Sequence 3, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; OPERATING SYSTEM: Version #1.25

APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 471..2848
US-08-537-361E-1

Query Match 60.68; Score 18.8; DB 3; Length 3319;
Best Local Similarity 76.78; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcacgaggacagaataaaactgtggaata 30
||||| ||||| ||||| |||||
Db 1298 TCACGAGGGCATAATTACAGGTTGAAGA 1327

RESULT 8
US-08-429-742-3
; Sequence 3, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; OPERATING SYSTEM: Version #1.25

APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3319 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 471..2848
US-08-537-361E-1

Query Match 60.68; Score 18.8; DB 3; Length 3319;
Best Local Similarity 76.78; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tcacgaggacagaataaaactgtggaata 30
||||| ||||| ||||| |||||
Db 1298 TCACGAGGGCATAATTACAGGTTGAAGA 1327

RESULT 8
US-08-429-742-3
; Sequence 3, Application US/08429742
; Patent No. 5686257
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Jacqueline
; APPLICANT: Bazan, J. Fernando
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN T CELL ANTIGENS AND
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; OPERATING SYSTEM: Version #1.25

APPLICANT: Heffron, Fred
APPLICANT: Nassif, Xavier
TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE:
```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (2526)
OTHER INFORMATION: n represents a, c, t or g.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2549)..(2552)
OTHER INFORMATION: n represents a, c, t or g.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2612)..(2614)
OTHER INFORMATION: n represents a, c, t or g.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2633)
OTHER INFORMATION: n represents a, c, t or g.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2920)
OTHER INFORMATION: n represents a, c, t or g.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3069)
OTHER INFORMATION: n represents a, c, t or g.
US-09-103-663-35

Query Match 58.7%; Score 18.2; DB 4; Length 4550;
Best Local Similarity 74.2%; Pred. No. 31;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 tcagcaggacagaataaaactgtggaatat 31
||||| ||| ||||| || || || ||
Db 4472 TTACGACAGACTGAATAAATCTATGATAAAT 4442

RESULT 10
US-08-781-891-79/c
Sequence 79, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781.891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 60906201enbun Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-79
Query Match 58.1%; Score 18; DB 3; Length 87350;
Best Local Similarity 80.8%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 6 agggacagagaataaaactgtggaatat 31
||||| ||| ||||| ||||| || || ||
Db 49985 AGGTACACAGATAAAACTGTAGCCTTT 49960
RESULT 11
US-08-325-553-1
Sequence 1, Application US/08325553
Patent No. 5538866
GENERAL INFORMATION:
APPLICANT: Israeli, Ron S.
APPLICANT: Heston, Warren D.W.
APPLICANT: Fair, William R.
TITLE OF INVENTION: THE PROSTATE-SPECIFIC MEMBRANE ANTIGEN
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325.553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/973.337A
FILING DATE: 05 NOV 1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1747/41426
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2653 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Carcinoma
IMMEDIATE SOURCE:
CLONE: Prostate-Specific Membrane Antigen
FEATURE:
NAME/KEY: CDS
LOCATION: 262..2511
US-08-325-553-1

RESULT 13
US-08-530-010-1/c
; Sequence 1, Application US/08530010
; Patent No. 5689055
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliott M.
; APPLICANT: Chang, Caren
; APPLICANT: Bleeker, Anthony B.
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,010
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/086,555
; FILING DATE: 01-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-57515/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-530-010-1

Query Match 56.8%; Score 17.6; DB 1; Length 3879;
Best Local Similarity 83.3%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 caggagacagataaaactgtgaa 28
|| ||||| ||||| |||||
DB 455 CAACACAGACAGAACTGTGAA 432

RESULT 14
US-08-484-101B-1/c
; Sequence 1, Application US/08484101B
; Patent No. 5824868
; GENERAL INFORMATION:
; APPLICANT: California Institute of Technology
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
; TITLE OF INVENTION: ETHYLENE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

Query Match 56.8%; Score 17.6; DB 1; Length 2653;
Best Local Similarity 83.3%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggacagataaaactgt 24
| ||||| ||||| |||||
DB 508 TTAGCAGGACAGACAAACTTT 531

RESULT 12
US-08-394-152A-1
; Sequence 1, Application US/08394152A
; Patent No. 5935818
; GENERAL INFORMATION:
; APPLICANT: Israeli, Ron S.
; APPLICANT: Heston, Warren D.W.
; APPLICANT: Fair, William R.
; TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,152A
; FILING DATE: 24-FEB-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41426-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2653 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Carcinoma
; IMMEDIATE SOURCE:
; CLONE: Prostate-Specific Membrane Antigen
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 262...2511
US-08-394-152A-1

Query Match 56.8%; Score 17.6; DB 2; Length 2653;
Best Local Similarity 83.3%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcagcaggacagataaaactgt 24
| ||||| ||||| |||||
DB 508 TTAGCAGGACAGACAAACTTT 531

Search completed: October 24, 2001, 10:00:14
Job time: 199 sec

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,101B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/
FILING DATE: 01-JUL-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/086,555
FILING DATE: 01-JUL-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-57515-2/RFT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3879 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-484-101B-1

Query Match 56.8%; Score 17.6; DB 1; Length 3879;
Best Local Similarity 83.3%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 cagggacagaaataaaactgtggaa 28
||| ||||| ||||| ||||| |||||
Db 455 CACACAGAGAAGAAACTGTGGAA 432

RESULT 15
US-09-173-914-5/c
Sequence 5; Application US/09173914
Patent No. 6171857
GENERAL INFORMATION:
APPLICANT: Hendrickson, Eric
TITLE OF INVENTION: A No. 6171857el Leucine Zipper, KARP-1 and
METHODS OF REGULATING DNA DEPENDENT PROTEIN KINASE ACTIVITY
FILE REFERENCE: B0877/7017/HK
CURRENT APPLICATION NUMBER: US/09/173,914
CURRENT FILING DATE: 1998-10-16
EARLIER APPLICATION NUMBER: 60/064,557
EARLIER FILING DATE: 1997-10-17
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 1094
TYPE: DNA
ORGANISM: Salmonella Typhimurium
US-09-173-914-5

Query Match 55.5%; Score 17.2; DB 4; Length 1094;
Best Local Similarity 73.3%; Pred. No. 64;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 tcacagcagcagacagacggttagttgattga 30
||||| ||||| ||||| ||||| |||||
Db 1074 TCAGCAGCGACACAGACGTTAGTTGATTGA 1045

us-09-462-955-1_copy_941_971.rni

Thu Oct 25 13:08:35 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:11 ; Search time 393.52 seconds
(without alignments)
673.344 Million cell updates/sec

Title: US-09-462-955-1_COPY_583_1004

Perfect score: 422
Sequence: 1 cttgaagtgtctgtcgat.....catctaaattacacccaatac 422

Scoring table: IDENTITY_NUC
Gapop 10.0'', Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.2	27.1	1096	20 AAV71833	BBTV DNA II clone
2	114.2	27.1	1096	21 AAA38950	Banana bunchy top
3	112.6	26.7	1091	20 AAV71834	BBTV DNA II clone
4	112.6	26.7	1091	21 AAA38947	Banana bunchy top
5	112.4	26.6	886	20 AAA34687	DNA sequence of BB
6	96.8	22.9	1106	20 AAV71832	BBTV DNA I clone (
7	95.4	22.6	1017	17 AAT13165	SCSV segment 6. S
8	95.2	22.6	1106	21 AAA38946	Banana bunchy top
9	88.4	20.9	1022	17 AAT13161	SCSV segment 2. S
10	80.4	19.1	593	20 AAA34686	DNA sequence of BB
11	62.8	14.9	1110	19 AAV24089	Banana bunchy top

12	59.6	14.1	982	18 AAV49405	Banana bunchy top
13	59.6	14.1	1110	19 AAV24086	Banana bunchy top
14	59.6	14.1	1111	19 AAV24077	Banana bunchy top
15	59.6	14.1	1111	19 AAV24084	Banana bunchy top
16	59.6	14.1	1111	19 AAV24087	Banana bunchy top
17	58.6	13.9	1103	19 AAV24091	Banana bunchy top
18	58.6	13.9	1105	19 AAV24093	Banana bunchy top
19	58	13.7	1109	19 AAV24088	Banana bunchy top
20	58	13.7	1111	19 AAV24090	Banana bunchy top
21	57	13.5	1104	19 AAV24092	Banana bunchy top
22	56.4	13.4	1111	19 AAV24085	Banana bunchy top
23	49.4	11.7	936	22 AAF58252	Oligonucleotide D1
24	49.4	11.7	936	22 AAF58253	Oligonucleotide D1
25	49.4	11.7	936	22 AAF58257	Oligonucleotide D1
26	49.4	11.7	936	22 AAF58259	Oligonucleotide D1
27	49.4	11.7	936	22 AAF58262	Oligonucleotide D2
28	49.4	11.7	938	22 AAF58255	Oligonucleotide D1
29	46	10.9	936	22 AAF58252	Oligonucleotide D1
30	46	10.9	936	22 AAF58254	Oligonucleotide D1
31	46	10.9	936	22 AAF58257	Oligonucleotide D1
32	46	10.9	936	22 AAF58259	Oligonucleotide D1
33	46	10.9	936	22 AAF58262	Oligonucleotide D2
34	46	10.9	938	22 AAF58255	Oligonucleotide D1
35	42	10.0	300	20 AAV71831	Subgenomic fragmen
36	42	10.0	300	21 AAA38949	Banana bunchy top
37	41	9.7	287	21 AAV71830	Subgenomic fragmen
38	41	9.7	287	21 AAA38948	Banana bunchy top
39	38	9.0	244	22 AAF58238	Oligonucleotide D1
40	38	9.0	244	22 AAF58238	Oligonucleotide D1
41	32.6	7.7	5053	19 AAV22816	Nucleotide sequenc
42	32.6	7.7	5055	19 AAV17236	DNA from a region
43	31.4	7.4	41599	21 AAZ35351	Cosmid including s
44	30.8	7.3	41599	21 AAZ35351	Cosmid including s
45	30.6	7.3	9345	21 AAZ39366	Mouse uromodulin p

ALIGNMENTS

RESULT 1
AAV71833
ID AAV71833 standard; DNA; 1096 BP.
XX
AC AAV71833;
XX
DE 10-FEB-1999 (first entry)
XX
DE BBTV DNA II clone (2-17) nucleotide sequence.
XX
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
XX
OS Banana bunchy top virus.
XX
FH Key Location/Qualifiers
FT TATA_signal 1..7
FT stem_loop /*tag= a
FT /*tag= b
FT CDS 70..928
FT /*tag= c
FT /*product= "ORF-V2 product"
FT /*transl_except= (pos:215..217, aa:Gly)
FT polyA_signal 533..538
FT /*tag= d
FT polyA_signal 799..804
FT DNA sequence of BB /*tag= e
FT BBTV DNA I clone (907..912
FT SCSV segment 6. S /*tag= f
FT Banana bunchy top 1030..1035
FT SCSV sequence 2. S /*tag= ge
XX
XX US5846705-A.

XX	23-AUG-2000	(first entry)
XX	Banana bunchy top virus related nucleotide sequence #5.	
XX	Banana bunchy top virus; BBTV; detection; ds.	
XX	Banana bunchy top virus.	
XX	TW360710-A.	
XX	11-JUN-1999.	
XX	30-JUN-1994; 94TW-0106105.	
XX	30-JUN-1994; 94TW-0106105.	
XX	(BIOT-) DEV CENT BIOTECHNOLOGY.	
XX	Wu R, You L, Song T;	
XX	WPI; 2000-316145/27.	
XX	Two circular single-stranded DNAs associated with banana bunchy top virus and detection of the virus.	
XX	Claim 2; Page 1; 7pp; Chinese.	
XX	The present invention describes two circular single-stranded DNAs associated with banana bunchy top virus (BBTV) and the encoded protein. Also described is a method of detecting the BBTV virus using the polymerase chain reaction (PCR). The present sequence represents a BBTV related nucleotide sequence from the present invention.	
XX	Sequence 1096 BP; 347 A; 231 C; 244 G; 274 T; 0 other;	

Query Match	27.1%	Score 114.2	DB 21	Length 1096
Best Local Similarity	57.5%	Pred. No. 2.3e-28		
Matches	225	Conservative	0	Mismatches 163; Indels 3; Gaps
QY	19	atcggagagccagcgagcagatcgacacatccctctggatatcgcgacgagcagcgaggagac	78	
Db	551	atcgagtggttctctgatgatcgagatcatctcgtgtatcgggtcccaacgagcgccaa	610	
QY	79	gggaagtcgcgtgttggccaaatatctcgagactcaagcccgactggtctacacatgtggt	138	
Db	611	ggaaagtcaccttcgacagatatactcattaaacccggatgggatatatacaacggt	670	
QY	139	ggaaacagaaagagcattgtaccagtcacatcgagggaccacaaacgaatttaacctc	198	
Db	671	ggaaagacgtcggtatgatgcacatcataacgatggatccgtgataatcattggattatt	730	
QY	199	gatgtaccocaggtgttaatttagagttattaaattatgcctgttagaagtggttaagac	258	
Db	731	gatatcccgagaagtcacagattatcgtattatggttatgagacaaattagaat	790	
QY	259	agggcattcagtcggaacaaatcacgaacc---ccttagttatctctgggttgaccatgtg	315	
Db	791	agagttttaataataacaaatcacgaacctgtgtgattagaagaatggacaaatgtc	850	
QY	316	catgtactcgtatttggcaatgctctgcctgattatttgaataatcagcgaggcagagaata	375	
Db	851	catgtaattgttatggcaaatgtgtgcctgtattgtgaaatttcagaagaatagaata	910	
QY	376	aaactgggaatatattaaagtattgtctatc	406	
Db	911	aaataaattaatctttgagaagaacgaacttc	941	

RESULT	2	3
AAA38950	AAA38950 standard; DNA; 1096 BP.	RESULT
ID	AAA38950	AAV71834
XX	AC	ID
AAA38950;		AAV71834 standard; DNA; 1091 BP.

Query Match	20.9%;	Score	88.4;	DB	17;	Length	1022;
Best Local Similarity	54.3%;	Pred. NO.	1.1e-19;				
Matches	201;	Conservative	0;	Mismatches	166;	Indels	3;
Gaps							
1;							
yy	25	gagccagcgacgatcgacaatcctcttgatgatacggcagcagacgagcagcgagggaag	84				
bb							
yy	559	gcggaaaccagatgatcgagatcatctgggtctatggtttcagacgaggaagaagaaag	618				
bb							
yy	85	tcggtgtttgccaaatatctcggaactcaagccgcagctggtctcacatatggtgtggaacc	144				
bb							
yy	619	acgagcttcgcgaagaaatt--aatcaggtatggtggtttttacagccgaggggaag	675				
bb							
yy	145	agaaagagacgtattgtaccagtatcatcgaggaccacaaacgaaatttaatcctcgatga	204				
bb							
yy	676	accagagcgtattatataatgtatgtctcaagcccgagaggaataatgcgtttgatgtt	735				
bb							
yy	205	cccagggttaattagatgtattaaattatgcctgttagaatgtgttaagacacaggcca	264				
bb							
yy	736	cccagggttcttcggagatgaactatcaggcgatggagatgttgaagacacagagtt	795				
bb							
yy	265	ttcagttcggacaaatacgaacccttagttatctttgggttcgaccatgtgcattactc	324				
bb							
yy	796	tttgcagtagacaaataaggcctgtagactctttagtagaagttagttcatttaatt	855				
bb							
yy	325	gtatttgcaatgtcctgcctgattatttgaaatacagcaggacagacagataaaactgtgg	384				
bb							
yy	856	gtgtttgccacgtggcacctgcacccacgcgcataagtgaggacagactgtgaattatc	915				
bb							
yy	385	aattattcaaa	394				
bb							

... 305

305

Thu Oct 25 13:08:15 2001

```

QY 126 ctacacatgtgglggaacacagaagagcgctattgtaccagttacatcgagagacccaaaaacg 185
Db 704 ttattctccagagagaaaatcatgtgatatgtagactgtataattacgaggatat--- 760
QY 186 aaatttaacctcgatgtaccaggtgtaatttttagagattttaaatatgccctgtttaga 245
Db 761 ---tgttatatttgatatcccaagatgcaagagagattatttaaatkatgggttatttaga 817
QY 246 atgtgttaagaacaggggcatttcagttcggacacaaatacgaaccccttagttatcttgggtt 305
Db 818 ggaatttaagaatggaataattcaaaagcggaataatgaaacccgttttgaagatag---t 874
QY 306 cgaccatgtgcactgactgtattgtccaatgtcctgcctgtgattatttgaaaaatcagcag 365
Db 875 agaatagtcgaagtcattgttaagtgcctaaacttccttcgcgaaggaagaatctttctga 934
QY 366 ggacagaataaaactg 381
Db 935 agatogaataaagtgtg 950

```

Search completed: October 24, 2001, 10:07:12
Job time: 617 sec

```
GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: October 24, 2001, 13:13:38 ; Search time 5479.82 Seconds
      (without alignments)
      727.963 Million cell updates/sec

Title: US-09-462-955-1_COPY_583_1004
Perfect score: 422
Sequence: 1 cttgaagtgtgtgtgcgat.....catctaaattacaccaatac 422

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estomi:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estrol:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*
```

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Plate:	720	row:	D	column:	3
Seq primer:	SP6				
Class:	BAC ends				
High quality sequence stop:	498				
Location/Qualifiers					
1..498					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone_lib="RPC1-11 Human Male BAC Library"					
/sex="male"					
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"					
BASE COUNT	171 a	107 c	87 g	113 t	20 others
ORIGIN					
Query Match		8.7%	Score 36.8;	DB 231;	Length 498;
Best Local Similarity		56.7%;	Pred. NO. 1.5;		
Matches	68;	Conservative	0;	Mismatches 52;	Indels 0; Gaps 0;
QY	250	gttaagaacagggcattcagttcggaacaatacgaacccttagttattcttggttcgac	309		
Db	271	gTCTGTGCCAGGGCAATCAGCAGCAAGCAAGAATAAAGGTTATTAACTAGGAAACAGAG	330		
QY	310	catgtgcattactcgtattttgccaattgctcctgcctgattatttgaataatcagcagggac	369		
Db	331	GAAGTCAAATTGTCCTTTTGGCAGATGACATGATTGTTTATTAGAAATCCCATCGAC	390		
RESULT	3				
AW818171		518 bp	mRNA	EST	17-MAY-2000
LOCUS	CMI-ST0277-161299-070-h05	ST0277	Homo sapiens	cdNA,	mRNA sequence.
DEFINITION	AW818171				
ACCESSION	AW818171.1	GI:7911165			
VERSION					
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	1 (bases 1 to 518)				
	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,				
	Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,				
	Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,				
	Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare				
	,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and				
	Simpson,A.J.				
	Shotgun sequencing of the human transcriptome with ORF expressed				
	sequence tags				
	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
JOURNAL	20202663				
MEDLINE					
COMMENT	Contact: Simpson A.J.G.				
	Ludwig Institute for Cancer Research				
	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,				
	Brazil				
	Tel: +55-11-2704922				
	Fax: +55-11-2707001				
	Email: asimpson@ludwig.org.br				
	This sequence was derived from the FAPESP/LICR Human Cancer Genome				
	Project. This entry can be seen in the following URL				
	(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CMI-ST0277-16				
	299-070-h05&t3=1999-12-16&t4=1)				
	Seq primer: puc 18 forward				
	High quality sequence start: 76				
	High quality sequence stop: 183.				
FEATURES	Location/Qualifiers				
source	1..518				
	organism="Homo sapiens"				

FEATURES	source
High quality sequence stop: 318.	
Location/Qualifiers	
1. .318	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="Plate=3001 Col=8 Row=0"	
/clone_lib="CIT Approved Human Genomic Sperm Library D"	
/sex="male"	
/note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in E-Coli DH10B"	
77 a	66 c 66 g 108 t 1 others
BASE COUNT	

1	218
2	666
3	278
4	726
5	338

```

ORIGIN

Query Match      8.4%; Score 35.6; DB 224; Length 318;
Best Local Similarity 54.6%; Pred. No. 3.2;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 250 gttaaagacaggcattcagttcgacaaatacgaacccttagttatctgggtcgac 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 GTTCTGCCAGGCAATCAGCAGGAGAGAAATAAAGGGTATTTTAATTAGGAAGAG 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 catgtcgatgactcgtatttgcgaattcctgcgtgattatttgaaatacagcaggac 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 GATTTCAAATTGCTCTGTTTCAGATGACATGACTGTATATCTAGAAAACCGATTGTC 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 agataaaac 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 TCAGTCCAAC 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
CONS016H0 1101 bp DNA GSS 26-JUL-1999
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN16B20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL106734
VERSION
AL106734.1 GI:5623558
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Plasmod Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage ;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
            Location/Qualifiers
source          1..1101
                /organism="Drosophila melanogaster"
                /plasmid="pBelobAC11"
                /db_xref="taxon:7227"
                /clone_lib="DrosBAC"
                /clone="BACN16B20"
                /note="end : 17"
BASE COUNT     260 a 123 c 159 g 217 t 342 others
ORIGIN

Query Match      8.4%; Score 35.6; DB 219; Length 1101;
Best Local Similarity 30.1%; Pred. No. 4.2;
Matches 59; Conservative 51; Mismatches 86; Indels 0; Gaps 0;

QY 218 tagagtattaaatgccctgttagaatgtgttaaacaacaggcattcagtcgaca 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 TAAATWADTWATWBWTAATTTGTTTTRMSAAAARTTWTGGGAKADWTTATNTRGR 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 278 ataagaaccccttagttatcttgggttcgccactgcatgctactcgtatttccaatg 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 DGGGGCAACAGKVTADTATATGRKAKSGAARGCWKTGTRTATATWGGDRKAGW 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 338 tctgcctgattatttgaaatcacagcaggacagaataaaactgtggaatttaaagta 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Thu Oct 25 13:08:18 2001

```

BASE COUNT      154 a      60 c      59 g      187 t
ORIGIN

Query Match      8.4%; Score 35.4; DB 104; Length 460;
Best Local Similarity 50.9%; Pred. No. 4;
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 223 tatttaattatgcctgttagaattgtttaagaacaggcgattcagttcggacataac 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 283 gaacccttagttatctgttggttcgaccatgtgcattgctactgtatttgccaattgctctg 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 AACCTCCATAATTATGTTGTTTAAAGCTCTATTATTAATCACTAGTATTTGAACAAACACAGCAAG 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 cctgattattgaaatcagcaggagacagaataaaactgtggaat 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 TATAATTTTGAATTCATCTAGTGAATTAATATTCATGTCCGAT 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
BE603735 955 bp mRNA EST 02-MAR-2001
LOCUS HVSMEH0087003f Hordeum vulgare 5-45 DAP spike EST library
DEFINITION HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0087003f,
mRNA sequence.
ACCESSION BE603735
VERSION BE603735.2 GI:13187835
KEYWORDS EST.
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare
REFERENCE 1 (bases 1 to 955)
AUTHORS Wing, R., Close, T. J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D. W., Main, D. and
Wood, T.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics
JOURNAL Unpublished (2000)
COMMENT On Aug 21, 2000 this sequence version replaced gi:9861296.
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCTCTACTAAAGG
High quality sequence stop: 931.
Location/Qualifiers
1. .955
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEH0087003f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
HVCNDA0009 (5 to 45 DAP)
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/notes="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      303 a      205 c      213 g      220 t
ORIGIN

Query Match      8.4%; Score 35.4; DB 137; Length 955;
Best Local Similarity 52.0%; Pred. No. 4.7;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 68 acgagagacgaggaagtcgctgttttgccaaatattctcgactcaagcccgactgttct 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

Db 585 ACAGGGGACTGGGAACCTGGGTTACCAAACTAATCGCTTCAACACATCCCTTTTC 644
Qy 128 acacatgtgttggaacgaagacgtattgtaccagtcacatcgaggagcccaaacgaa 187
Db 645 ACAGTGGCGTATACAAAAAGGCGGAGGATTCCTCCCTTCCAAAAGTGCCCAATCTGAA 704
Qy 188 attaatcctcatgtaccagggtgtaatt 217
Db 705 GGGGAATGGGAATGGAGACGCTAATATT 734

RESULT 9
AZ401897/c
LOCUS
DEFINITION
  1M0168N22R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
  clone UUGCLM0168N22 R, DNA sequence.
VERSION
  AZ401897
KEYWORDS
  GSS.
SOURCE
  house mouse.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 345)
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  ,M., Rose,M., R., Stokes,R., Tingey,A., von Niederhausen,A.
  and Wright,D., Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0168 row: N column: 22
  Seq primer: CACACAGGAACACGCTATGACC
  Class: plasmid ends
  High quality sequence stop: 345.
  Location/Qualifiers
    1..345
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGCLM0168N22"
      /sex="Male"
      /lib_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
      /notes="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnaref/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA
      was blunt end-repaired with T4 DNA polymerase and T4
      polynucleotide kinase. Adaptor oligonucleotides were
      ligated to the blunt ends in high molar excess. The
      adapted DNA was purified and size-selected for a 9.5 to
      10.5 kb range using preparative agarose gel
      electrophoresis. Vector DNA was prepared from a derivative
      of pWD42 (g114732114/gb1AF129072.1), a copy-number
      inducible derivative of plasmid R1. The vector was ligated
      with adaptors complementary to the insert adaptors and
      purified. The sheared, adapted mouse DNA was annealed to
      adapted vector DNA, and transformed into
      chemically-competent E. coli XL10-Gold (Stratagene) cells
      and selected for ampicillin resistance."
FEATURES
  source
    1..345
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGCLM0168N22"
      /sex="Male"
      /lib_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
      /notes="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnaref/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA
      was blunt end-repaired with T4 DNA polymerase and T4
      polynucleotide kinase. Adaptor oligonucleotides were
      ligated to the blunt ends in high molar excess. The
      adapted DNA was purified and size-selected for a 9.5 to
      10.5 kb range using preparative agarose gel
      electrophoresis. Vector DNA was prepared from a derivative
      of pWD42 (g114732114/gb1AF129072.1), a copy-number
      inducible derivative of plasmid R1. The vector was ligated
      with adaptors complementary to the insert adaptors and
      purified. The sheared, adapted mouse DNA was annealed to
      adapted vector DNA, and transformed into
      chemically-competent E. coli XL10-Gold (Stratagene) cells
      and selected for ampicillin resistance."

```

```

BASE COUNT 104 a 77 c 58 g 106 t
ORIGIN

Query Match
Best Local Similarity 8.3%; Score 35.2; DB 243; Length 345;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 216 tttagatatttaattatgccctgttagaattgtttaagaacagggcatttcagtcga 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 TTTCAGAGCTTAAATGGGACGCCCTCTGAATGAGGTATAATATGCTATTACGTCATGA 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 276 caaatagaaccccttagttatcttgggttcgaccatgtgcattgactctgatttcgcaa 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 AACCTGAGGACATCTTTGTGGTGGTAGTTGACTGCTGACTTGCCTTGCCTGCTGTGTCAA 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
AQ746046
LOCUS
DEFINITION
  HS_2277_A2_F11_J7C CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=2277 Col=22 Row=K, DNA sequence.
ACCESSION
  AQ746046
VERSION
  AQ746046.1 GI:55233568
KEYWORDS
  GSS.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 848)
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  99380589
  Contact: Mahairas GG, Wallace JC, Hood L
  High throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3887
  Email: jwallace@u.washington.edu
  Clones may be purchased from Research Genetics (info@resgen.com).
  BAC end Web Server: http://www.htsc.washington.edu
  Plate: 2277 row: K column: 22
  Seq primer: T7
  Class: BAC ends
  High quality sequence stop: 848.
  Location/Qualifiers
    1..848
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="Plate=2277 Col=22 Row=K"
      /clone_lib="CIT Approved Human Genomic Sperm Library D"
      /sex="male"
      /notes="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
      E-Coli DH10B"
BASE COUNT 228 a 136 c 117 g 144 t 223 others
ORIGIN

Query Match
Best Local Similarity 8.3%; Score 35.2; DB 233; Length 848;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 250 gttagaacagggcatttcgagacaaatcgaaccccttagttatcttgggttcgac 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 GTTCTGCCAGGCCAATCAGTAGGAGAAAGAAATTAACCGGTATTCAATAGGAAAGAG 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 310 catgtcatgtactgtatgtcccaatgtcctgcctgatttgaataatcagcaggac 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

1

Thu Oct 25 13:08:18 2001

Query Match 8.0%; Score 33.8; DB 219; Length 604;
Best Local Similarity 62.4%; Pred. No. 13;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 178 ccaaaacgaaatttaactcgcgtgtaccaggtgttaatttagagttatttaaattatgcc 237
Db 311 CTAAACAAATTGAAGTGCACCTCTTAACACAGGTGAAATTTATAGTGTGCAAAATTATATA 370
QY 238 ctgttagaatgtgttaagaacagg 262
Db 371 CTAATAAAACCTGTTTAAAAAAAGG 395

Search completed: October 24, 2001, 13:13:40
Job time: 11805 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:14 ; Search time 180.6 Seconds
(without alignments)
442.355 Million cell updates/sec

Title: US-09-462-955-1_COPY_583_1004

Perfect score: 422
Sequence: 1 cttgaagtctgtctgcgat.....catctaaattacaccaatac 422

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.2	27.1	1096	2	US-08-418-071-4
2	112.6	26.7	1091	2	US-08-418-071-5
3	96.8	22.9	1106	2	US-08-418-071-3
4	95.4	22.6	1017	4	US-08-793-634B-6
5	88.4	20.9	1022	4	US-08-793-634B-2
6	62.8	14.9	1110	1	US-08-202-186-14
7	59.6	14.1	982	3	US-08-973-068-14
8	59.6	14.1	1110	1	US-08-202-186-11
9	59.6	14.1	1111	1	US-08-202-186-9
10	59.6	14.1	1111	1	US-08-202-186-12
11	59.6	14.1	1111	1	US-08-202-186-24
12	58.6	13.9	1103	1	US-08-202-186-16
13	58.6	13.9	1105	1	US-08-202-186-18
14	58	13.7	1109	1	US-08-202-186-13
15	58	13.7	1111	1	US-08-202-186-15
16	57	13.5	1104	1	US-08-202-186-17
17	56.4	13.4	1111	1	US-08-202-186-10
18	42	10.0	300	2	US-08-418-071-2
19	41	9.7	287	2	US-08-418-071-1
20	29.4	7.0	8920	2	US-08-446-855A-1
21	29.4	7.0	8920	4	US-09-150-741-1
22	29	6.9	9468	1	US-08-325-547-10
23	28.8	6.8	10607	1	US-08-078-090-3
24	28.6	6.8	2899	2	US-08-624-581-2
25	28.6	6.8	2917	2	US-08-624-581-3
26	28.6	6.8	2959	2	US-08-624-581-1
27	28.4	6.7	1380	4	US-09-499-505-4

c

c	28	28.4	6.7	1430	4	US-09-499-505-3	Sequence 3, Appli
c	29	28	6.6	2483	1	US-08-526-964-1	Sequence 1, Appli
c	30	28	6.6	2483	3	US-08-946-617-1	Sequence 5, Appli
c	31	28	6.6	2483	3	US-09-031-897-1	Sequence 1, Appli
c	32	28	6.6	19932	2	US-08-477-451-25	Sequence 25, Appli
c	33	28	6.6	28958	1	US-08-258-261B-6	Sequence 6, Appli
c	34	28	6.6	28958	1	US-08-456-837-6	Sequence 6, Appli
c	35	28	6.6	28958	1	US-08-457-342-6	Sequence 6, Appli
c	36	28	6.6	28958	1	US-08-457-646A-6	Sequence 6, Appli
c	37	28	6.6	28958	1	US-08-458-076A-6	Sequence 6, Appli
c	38	28	6.6	28958	1	US-08-764-233A-4	Sequence 4, Appli
c	39	28	6.6	28958	1	US-08-457-335A-6	Sequence 6, Appli
c	40	28	6.6	28958	1	US-08-729-214-6	Sequence 6, Appli
c	41	28	6.6	28958	3	US-09-028-934-6	Sequence 6, Appli
c	42	28	6.6	49377	1	US-08-764-233A-1	Sequence 1, Appli
c	43	27.8	6.6	2570	2	US-09-056-075-2	Sequence 2, Appli
c	44	27.6	6.5	6476	4	US-09-127-670-5	Sequence 5, Appli
c	45	27.4	6.5	840	5	PCT-US91-08177-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705

GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh

; APPLICANT: You, Li-Ru

; APPLICANT: Soong, Tai-Seng

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE

; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: United States

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 5.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/418,071

; FILING DATE: 06-APR-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley Jr., James F.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: DCB-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 596-9000

; TELEFAX: (212) 596-9090

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1096 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: circular

; MOLECULE TYPE: genomic DNA

; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Banana Bunchy Top Virus (BBTV)

; US-08-418-071-4

Query Match	20.9%	Score 88.4	DB 4	Length 1022	
Best Local Similarity	54.3%	Pred. No. 1.4e-20			
Matches 201	Conservative	0	Mismatches 166	Indels 3	Gaps 1
yy	25	gagccagcgagcagatcgacaaatcctctgatatcgagcagagacgagcgagcgagcggaag	84		
bb	559	GCGGAAACCATGATCGGAGCTATCATCTGGTCTATGTTTCAGACGAGGAGAGAAAG	618		
yy	85	tcctgtttgccaaatatctcggaactcaagcccgactggtttctcacatgtgtgtgaacc	144		
bb	619	ACGAGCTTCGGAAGGAATT--AATCAGGTATGGATGTTTTTATACGCGGAGGGGAAG	675		
yy	145	agaaagacgtattgtaccagtatcatcgaggaccccaaaacgaatttaatcctcgtatgta	204		
bb	676	ACCAGAGCGATATATATATGTATGTCTCAGACCCAGAGGAATATTCGGTTGATGTT	735		
yy	205	cccaggtgtaatttagagattttaaatatgccttcttagaatgtgttaagacaacgggca	264		
bb	736	CCCAGGTGTTCTTCGGAGATGATGAACATAICAGCGCATGGAGATGTTGAAGAACAGAGTT	795		
yy	265	ttcagttcggacaaatacgaaccccttagtactctcgggttcgacatgtgcagtactc	324		
bb	796	TTTGCAAGTACAAAATATAGCCCTGTAGATCTTTGTATTAGGAAGCTAGTTCATTTAATT	855		
yy	325	gtatttgcgaatgtcctgcctgattatttgaataatcagcagggcagagaataaaactgtgg	384		
bb	856	GTGTTGCCACAGTGGCACCCTGACCCACCGCATAGTGAGGACACAGACTGTGTAATTATC	915		
yy	385	aatatttaa	394		
bb	916	AATTGTTGAA	925		

US-08-973-068-28
: Sequence 28, Application US/08973068

Query Match	14.9%;	Score 62.8;	DB 1;	Length 1110;
Best Local Similarity	53.2%;	Fried. No. 7.7e-12;		
Matches 183;	Conservative	0; Mismatches 152;	Indels	Gaps
y	38	atcgacaatcctctdgatatcgacagacgaggaggaogggaaagccgtgtgttgcca	97	
b	615	ATCGAAGATTATTTCGGGTTTACGGCCCAATTCGGAGAGAAAGAACAACGTATGC	674	
y	98	aatatctcggaactcaagcccgcactggtctcacacatgtgttggaaaccagaaggcgat	157	
b	675	AACATCTTAATGACAGACGAAGATCGCTTTTATCTCCAGGAGAAAATCATTTGGATA	734	
y	158	tgtaccagtagcatcgaggaccocaaaaatttaalcctcagtatgccaggtgaatt	217	
b	735	GTAGACTGTATAATTATGAGGATAT-----TGTTATATTTCATATPCCCTAGATC	788	
y	218	tagagtatttaaattatgccctgtagaattgtttaagacacagggcattcagttcggaca	277	
b	789	ACGAATTTTAAATTATGCTTTTATAGAGAAATTTAAGAATGGAATAATCAAAAGCGGA	848	
y	278	aatacgaaccccttagtatacttgtgggttcgaccatgtcagtcactcgtatttgcgaatg	337	
b	849	AATATGAACCGTTTTTGAAGATTG---TAGAATATTCGAAGHCATGTAANTGGCTACT	905	
y	338	tcttcacctgattatttgaataatcagcagggcagagataaaaactg	381	
b	949	TGTTTTCACCAACCAATGTTTTCACCAATGTTTTCACCAATGTTTTCACCAATGTTT	949	

RESULT

RESULTS / DISCUSSION - 973-068-28

US-08-973-068-28
: sequence 28, Application US/08973068


```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-08/202,186
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: JEFFERY, Donald D.
REGISTRATION NUMBER: 19,980
REFERENCE/DOCKET NUMBER: 71611/102 FIVE
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1110 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-11

```

Query Match 14.1%; Score 59.6; DB 1; Length 1110;
Best Local Similarity 51.3%; Pred. No. 9,5-e-11;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps

Query Match	14.1%	Score 59.6;	DB 1;	Length 1110;
Best Local Similarity	51.3%;	Pred. NO. 9.5e-11;		
Matches 193;	Conservative 0;	Mismatches 174;	Indels 9;	Gaps
QY	6	adgtgctgctgcgatcgagagcagcgagcgagcagatcgacaaatcctctgatatcgcgacg	65	
Db	584	AGAGTGTGAGAAATATCATGCGGAGCGCATGTGCATGGAGAAATATTTGGGTCTATGGACC	643	
QY	66	agacgaggagacggaagtcctgtttgccaaatattctcgactcaagcccgactggtt	125	
Db	644	AAATGGAGGAGAGGAAGACAACGATGCAANAACATCTATATGAGAGCAGAAATGCGTT	703	
QY	126	ctacacatgtgttggaaccaagaagcagcattgtaccagtcatactcgaggaccacaaacg	185	
Db	704	TTATTTCCAGGAGGAAATCATTTGGATATATGTAGACTGTATAATTACGAGGATAT--	760	
QY	186	aaatttaactctcgatgaccagggtgtaatttagagtattaaattatccctgtttaga	245	
Db	761	---TGTTATATTGATATTCCAAGATGCCAAGAGAGATTTATTTAAATTATGGTATTATAGA	817	
QY	246	atgtgtttaagacagggccattcagttcggacaaataacgaccccttagttatcttgggtt	305	
Db	818	GGAATTTAAGAAATGGAATTAATTCAAGCGGGAAATATCAACCCGTTTGAAGATAG---T	874	
QY	306	cgacatgtgcagtactcgatttgcgaatgtcctgcgtgattattgaaaatcagcag	365	
Db	875	AGAATATGTCGAAGTCAATTTGATGGCTAACCTTCCTCCGAAGGAAGAAATCTTTTCTGA	934	
QY	366	ggacagaataaaactg	381	
Db	935	AGATCGAAATAAAGTTG	950	

```

RESULT          9
US-08-202-186-9
; Sequence 9, Application US/08202186
; Patent No. 5756708
;
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.

```


APPLICANT: DALE, James L.
 APPLICANT: HARDING, Robert M.
 TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
 NUMBER OF SEQUENCES: 60
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W.
 CITY: Washington, D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/202.186
 FILING DATE: 24-FEB-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: JEFFERY, Donald D.
 REGISTRATION NUMBER: 19,980
 REFERENCE/DOCKET NUMBER: 71611/102 FTKE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202 672 5300
 TELEFAX: 202 672 5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1105 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: circular
 IS-08-202-186-18

RESULT 15
US-08-202-186-15
; Sequence 15, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:

RESULT 14
US-08-202-186-13
: Sequence 13, Application US/08202186
: Patent No. 5756708
: GENERAL INFORMATION:
: APPLICANT: KARAN, Mirko
: APPLICANT: BURNS, Thomas M.
: APPLICANT: DALE, James L.

Job time: 201 sec

Query Match	13.7%	Score 58;	DB 1;	Length 1111;
Best Local Similarity	51.1%;	Pred. No. 3.3e-10;		
Matches 192;	Conservative	0;	Mismatches 175;	Indels
				Gaps
QY	6	agtgtgltcgtcgatcgagaagccagcgagcgtatcgacaatectctgdatcgcgacg	65	
Db	584	AGAGGTGGAGAAATATCATGGCGCAGCCATGTCATPCGGAGAAATAATTGGTGCTATGGCCC	643	
QY	66	agacggagagacgaggaaagtcgtgttgccaaatatctcgagctcaagcccgactggtt	125	
Db	644	AAATGGAGGAGAGAAAGACAACGATATGCAAAACATCTAATCAACAGCAGAAATCGCGTT	703	
QY	126	ctacacatgtgtggaaccgaaggaagcgtatgtaccggtacatcgaggaccacaaacg	185	
Db	704	TTATTCCTCAGGAGGAAACATATGGATATATGATAGACTGTATAATTCACGAGGATAT---	760	
QY	186	aaatttaaactcgtatgaccagggtgtaattlagaagtattaaattatgccctgttaga	245	
Db	761	---TGTTATATTGATATTCAGATGTCGAAGAGGATATTTAAATTTATGGGTATTAGA	817	
QY	246	atgtgttaagaacaggcattcagtcgttcggaacaatacacacccttagttatcttgggtt	305	
Db	818	GGAAATTTAAGAAATGGAATAATTCAAACGCGGAATATGAACCCGTTTTCGAAGATAG--T	874	
QY	306	cgaccatgtgcattgactcgtattttgccaatgctcgtcgtattatttggaaaaatcagcag	365	
Db	875	AGAATATCTCCAGCTCATGTAAATGGCTAACTTCCTCCGAAGGAAGGAATCTTTCCTGA	934	
QY	366	ggacagagaataaaaactg	381	
Db	935	AGATCGAATATAGTTG	950	

us-09-462-955-1_copy_583_1004.rni

Thu Oct 25 13:08:17 2001

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	120.6	14.5	1096	20	AAV17833	BBTV DNA II clone
2	120.6	14.5	1096	21	AAA38950	Banana bunchy top
3	117	14.0	1091	20	AAV17834	BBTV DNA II clone
4	117	14.0	1091	21	AAA38947	Banana bunchy top
5	115.6	13.9	886	20	AAA34687	DNA sequence of BBTV
6	102.2	12.3	1106	20	AAV17832	BBTV DNA I clone
7	100.6	12.1	1106	21	AAA38946	Banana bunchy top
8	99.2	11.9	1017	17	AAT13165	SCSV segment 6
9	91.2	10.9	1022	17	AAT13161	SCSV segment 2
10	81.2	9.7	593	20	AAA34686	DNA sequence of BBTV
11	62.8	7.5	1110	19	AAV24089	Banana bunchy top

ALIGNMENTS

RESULT 1

AAV71833
ID AAV71833 standard; DNA; 1096 BP.

XX AAV71833;

XX
DT 10-FEB-1999 (first entry)

BETV DNA II clone (2-17) nucleotide sequence.

Banana bunchy top virus; BBTv; PCR technique; plant; tissue; banana;
Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic

OS Banana bunchy top virus.

XX	FH	Key	Location/Qualifiers
----	----	-----	---------------------

```
FT TATA_signal 1.7 /*tag=
```

```
stem_loop 8..38 /#+30=
```

ST	CDS	70..928
7	ca9	

```

/*tag= C
/product= "ORF-v2 product"
/transl_except= (pos:215..217, aa:Gly)

```

```
polya_signal 333..338
/*tag=
```

```
polyA_signal 799..804
/*tag=
```

polyA_signal 907..912

polyA_signal

XX
1
/ Laq=

US5846705-A.

Thu Oct 25 13:08:03 2001

PF 30-JUN-1994; 94TW-0106105.
 XX 30-JUN-1994; 94TW-0106105.
 XX (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX Wu R, You L, Song T;
 XX WPI; 2000-316145/27.
 DR Two circular single-stranded DNAs associated with banana bunchy top
 PT virus and detection of the virus -
 XX Claim 2; Page 1; 7pp; Chinese.
 XX The present invention describes two circular single-stranded DNAs
 CC associated with banana bunchy top virus (BBTV) and the encoded proteins.
 CC Also described is a method of detecting the BBTV virus using the
 CC polymerase chain reaction (PCR). The present sequence represents a
 CC BBTV related nucleotide sequence from the present invention.
 XX Sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;
 SQ

Query Match 14.0%; Score 117; DB 21; Length 1091;
 Best Local Similarity 52.2%; Pred. No. 2.8e-25;
 Matches 314; Conservative 0; Mismatches 275; Indels 12; Gaps 2;

QY 9 ctgagcagcggagtcctcgcactcctcgtgagtcacaaaggccacgattggcccaacgattt 68
 DB 345 ctggaactgggtactcctcgtggttctcgttcgaagaagcgaagcttctcgatagattc 404
 QY 69 gctgaggaacacctgaaactccgcctggaagaccacagcggataccggaagatcggttga 128
 DB 405 agagagagccctgaagaattgaagatggagcagcatcccaagatcgagatgcttgga 464
 QY 129 cacggagcttcgtgaaatggacaaagatggccgctgaaatccgttcccttccattccat 188
 DB 465 gtggaatcaattaaagatgccgaataattccgaatgggttcagaaacta----- 515
 QY 189 cacaaatggcagctggaatgtctgcgtcggagagccagcggagcagcagcagcagcagc 248
 DB 516 aaagaatggcaaaataaataatcaacacatcgaaaggttctcgtgagcgaagatc 575
 QY 249 ctctgatatcgagcagagagagagagcgggaagtcctcgtgttgcacaaatattctcga 308
 DB 576 atctgggtatcacgtctcgtcggagcggaggaagtcacaccttcgcaagatatctatca 635
 QY 309 ctcaagccgactggttctacacatgtggtggaacacagaaagcagcgtattgtaccagtac 368
 DB 636 ttaaaacctggatgggatatatacaacggtggaagacgctcgatgatgcacatcata 695
 QY 369 atcgagagacccaaacaaatttaactctcgtatgacccaggtgtaatttagatattta 428
 DB 696 acgatgactcgtgataatcattgattgattgattcctccaggaagtcattcagattatctg 755
 QY 429 aattatccctgttagaattgttgaagacagggcattcagttcggacaaatcacgaacc- 487
 DB 756 aattatggcgtatagacaaataaagaatagatttttaataatacaaaatacgaacca 815
 QY 488 --cctagtattcttgggttcgaccatgtgcatgtactcgtatttgcgaatgtcctgct 545
 DB 816 tgtgtgattagaagaatggacaaaatgtccatgtattgttattgcaaatgtgtgct 875
 QY 546 gattattgaaatcagcagggacagaaataaactgtggaattttaaagtattgttgc 605
 DB 876 gattattgaaatcagcagggacagaaataaactgtggaattttaaagtattgttgc 935
 QY 606 c 606
 DB 936 c 936

RESULT 5
 AAX34687
 ID AAX34687 standard; DNA; 886 BP.
 XX
 AC AAX34687;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE DNA sequence of BBTV S2I and S2L promoter fragments.
 XX
 KW Promoter; Banana Bunchy Top Virus; BBTV; gene transcription; ubil;
 KW polyubiquitin 1; banana; ss.
 XX
 OS Banana bunchy top virus.
 XX
 PN WO9915646-A1.
 XX
 PD 01-APR-1999.
 XX
 PF 21-SEP-1998; 98WO-AU00786.
 XX
 PR 30-JUN-1998; 98AU-0004423.
 PR 19-SEP-1997; 97AU-0009339.
 XX
 XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 PA
 XX Becker DK, Dale JL, Dugdale B, Harding RM, Hermann SR;
 PI
 XX WPI; 1999-254706/21.
 DR
 XX Banana Bunchy Top Virus promoter constructs
 PT
 XX Examples; Fig 23; 84pp; English.
 PS
 XX The invention relates to DNA promoter sequences derived from components
 CC of Banana Bunchy Top Virus (BBTV). The DNA molecule includes a promoter
 CC sequence derived from an untranslated portion of any one BBTV components
 CC and is adaptable for promoting transcription of a cloned gene in a plant
 CC cell. The promoter, included in a DNA chimeric vector, is useful for the
 CC expression of a gene in a plant cell. The inclusion of the polyubiquitin
 CC 1 (ubil) intron sequence into BBTV promoter constructs enhances promoter
 CC activity. The present sequence represents the DNA sequence of BBTV S2I
 CC promoter and S2L promoter fragments.
 XX
 SQ Sequence 886 BP; 294 A; 171 C; 199 G; 222 T; 0 other;

Query Match 13.9%; Score 115.6; DB 20; Length 886;
 Best Local Similarity 52.4%; Pred. No. 7.5e-29;
 Matches 309; Conservative 0; Mismatches 269; Indels 12; Gaps 2;

QY 7 ttctcgagcagcggagtcctcgcactcgtcctcgagtcacaaaggccacgattggcccaacgat 66
 DB 69 ttctgaaactgggtactcctcgtggttctcgttcgaagaagcgaagcttctcgagagat 128
 QY 67 ttgctaaggaacactgatgaactccgctggaagacccagcggataccggaagatgcgttg 126
 DB 129 tcagagagagccctggaagaattgaagatggagatccatcccaagtatcgagatgcttgg 188
 QY 127 tacacggagcttcgttggatggacaagatggcgcgtgaaataccgttccattccat 186
 DB 189 cagtggaaatcgttgaacaaatgctaggaaaaattctgaatgggttc-----atgaac 239
 QY 187 atcaaatggcagcttgagtgctgctgcgtcgagagcagcggagcagcagcagcagcagc 246
 DB 240 taagagaatggcaaaataaataattcaacacatcgaaaggttctcgatgatcgagta 299
 QY 247 tcctcttgatgatcgagcagcggagagcggagcggagcggagcggagcggagcggagcggag 306
 DB 300 tcactctgggtatcaggtcccaacgggaggaaggaagcaaccccttcgcaagatatcat 359
 QY 307 gactcaagcccgactggttctacacatgtgttggaaccaggaagacgctatgtaccagt 366

Db 360 cattaaaaactggatggggatatacaacgggtggaagagctggatgatgatcaatca 419
 QY 367 acatcgagaccacaaacgaaatttaattcccgatgtaccaggtgtaatttagagtatt 426
 Db 420 taacgatgactcgataataatcattggattatgatcccaagaagtcattcgattatc 479
 QY 427 taattatgcccgttgaattgttgaagaacagggcattcagttcggacaaatacgaac 486
 Db 480 tgaattatgggttatagaacaaataagaataagagttttaataatacaaaatacgaac 539
 QY 487 c---ccttagttatcttgggttcgacacctgtgcatgtactctgtatttgcaatgtcctgc 543
 Db 540 catgtgtgattagaagaagatggacaaatgtccatgttaattgttatggcaaatgtgttc 599
 QY 544 ctgattattgaaatcagcagggacagagaataaaactgtggaaattttaa 593
 Db 600 ctgattattgaaatttcgaagatagaataaaaaataatttaattgttga 649
 RESULT 6
 AAV71832
 XX AC AAV71832 standard; DNA; 1106 BP.
 XX DT 10-FEB-1999 (first entry)
 XX DE BBTV DNA I clone (7-4-2) nucleotide sequence.
 XX KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
 XX KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
 XX OS Banana bunchy top virus.
 XX FH Key
 FT TATA_signal Location/Qualifiers
 FT 1..8
 FT /*tag= a
 FT stem_loop 9..38
 FT /*tag= b
 FT CDS 62..922
 FT /*tag= c
 FT /product= "ORF-V2 product"
 FT /transl_except= (pos:335..337, aa:Gly)
 FT /transl_except= (pos:518..520, aa:Ser)
 FT polyA_signal 380..385
 FT /*tag= d
 FT polyA_signal 901..906
 FT /*tag= e
 XX US5846705-A.
 XX PD 08-DEC-1998.
 XX PF 06-APR-1995; 95US-0418071.
 XX PR 06-APR-1995; 95US-0418071.
 XX PA (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX PI Soong T, Wu R, You L;
 XX WPI; 1999-059037/05.
 XX P-PSDB; AAW87459.
 XX Nucleic acids having banana bunchy top virus component sequences -
 PT used to design primers for use in polymerase chain reaction
 PT detection of the virus
 XX Claim 1; Fig 11A-B; 27pp; English.
 XX This represents the nucleotide sequence of a banana bunchy top virus
 CC (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The
 CC invention provides nucleic acid sequences associated with BBTV that can

CC be used in a PCR technique for detecting BBTV. The nucleic acid
 CC sequences (AAV71830 to AAV71833) are used as the basis for the
 CC construction of PCR primers, to detect BBTV infection. The PCR technique
 CC is used for detecting BBTV in plant tissues (preferably banana,
 CC especially Musa species). The virus, one of the most important banana
 CC species viruses, causes phloem damage and is transmitted by aphids. PCR
 CC detection gives accurate, reliable and specific determination of absence
 CC or presence of the virus.
 XX Sequence 1106 BP; 335 A; 225 C; 257 G; 289 T; 0 other;
 Query Match 12.3%; Score 102.2; DB 20; Length 1106;
 Best Local Similarity 53.3%; Pred. No. 3e-24;
 Matches 215; Conservative 0; Mismatches 188; Indels 0; Gaps 0;
 QY 195 tggcagcttgaagtgtctgtcgatcgagagccagcgagatcgcaatccctctg 254
 Db 524 tggcagattcaattgacggaggcaattgacgaggaacccgagatcggaagcatctgg 583
 QY 255 atatcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 314
 Db 584 gtctatggtctcttattgtaattgaggttaatacaacatcgcaagtcactaatacgaag 643
 QY 315 ccgactgttctacacatgtgttgaaccagaaagcagcagcagcagcagcagcagcag 374
 Db 644 gattgttctacaccagggttgaagaggaatacttcttctctctatgtgacgaa 703
 QY 375 gacccaaacgaaatttaattcctcgtatgtaccaggtgtaatttagatttaattat 434
 Db 704 ggatctgacaaacatagatttgaattcctcgttgaattcagcagcagcagcagcag 763
 QY 435 gcccttttagaattgttgaagacagggcagcagcagcagcagcagcagcagcagcagcag 494
 Db 764 gatgtaatagaggcattaaaggatagggttatagagagtactaaatacaaacccataaag 823
 QY 495 tatcttgggttcgaccatgtgcattactcgtattgttgcgaatgtcctgctgattatt 554
 Db 824 atagtgaattaggttaaatatcatgtatcgtcatggtcgaatttcctgctgactctgt 883
 QY 555 aaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 597
 Db 884 aaatctccgaagatcgaaataaaatcatttattgctgaagaa 926
 RESULT 7
 AAA38946
 ID AAA38946 standard; DNA; 1106 BP.
 XX AC AAA38946;
 XX DT 23-AUG-2000 (first entry)
 XX DE Banana bunchy top virus related nucleotide sequence #1.
 XX KW Banana bunchy top virus; BBTV; detection; ds.
 XX OS Banana bunchy top virus.
 XX TW360710-A.
 XX PD 11-JUN-1999.
 XX PF 30-JUN-1994; 94TW-0106105.
 XX PR 30-JUN-1994; 94TW-0106105.
 XX PA (BIOT-) DEV CENT BIOTECHNOLOGY.
 XX PI Wu R, You L, Song T;
 XX WPI; 2000-316145/27.

PR 07-NOV-1994;

XXX

Qy	386	aaattaatctcgtatgccagggtgaatttagagttattaaaattgacctgttga	445
Db	230	---tggtatttgtattccaagtgcgaagaggattttaataattagggttataga	286
Qy	446	atgtgttaagaacaggccattctgttcggacaatatcgacccttagttatctgggtt	505
Db	287	ggaattcaagaatggataaattcaaagcgggaatatgaaccggtttgaagatag--t	343
Qy	506	cgaactgtcatgtactcgtatttggcaatgtctcgtcgtattattgaaatacagcag	565
Db	344	agaatagtgcgaagtcatgtgaatggctaactctccgaagggaaggaatctttctga	403
Qy	566	ggacagaataaaaactg	581
Db	404	agatcgaataaaattg	419

RESULT 13

AAV24086
ID AAV24086 standard; DNA; 1110 BP.
XX
XX AC AAV24086;
XX
XX DT 11-AUG-1998 (first entry)
XX
XX DE Banana bunchy top virus component 1 DNA sequence.
XX
XX KW BBTv; probe; diagnostic primer; component 1; ss.
XX
XX OS Banana bunchy top virus.
XX
XX PN US5756708-A.
XX
XX PD 26-MAY-1998.
XX
XX PF 24-FEB-1994; 94US-0202186.
XX
XX PR 24-FEB-1994; 94US-0202186.
XX
XX PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
XX
XX PI Burns TM, Dale JL, Harding RM, Karan M;
XX
XX DR WPI; 1998-321636/28.
XX
XX PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers and for producing virus-resistant plants
XX
XX PS Disclosure; Column 19-22; 59pp; English.
XX
XX CC This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plants or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
XX enhancers or termination signals.
XX
XX SQ Sequence 1110 BP; 366 A; 191 C; 276 G; 277 T; 0 other;

	Query Match	7.2%	Score 59.6;	DB 19;	Length 1110;
	Best Local Similarity	51.3%;	Pred. No. 8.e-10;		
	Matches 193;	Conservative 0;	Mismatches 174;	Indels 9;	Gaps 2;
QY	206	aTgCtGtCTcgcATcgAgagAcacgAgcgAGctgcAcaaatcTctGgatcgGacg	265		
Db	584	agggTggagaatacAtcggcgcgcctcTcatcgagaaTaattggTgtctaTggacc	643		
QY	266	agacggaggagAcgggaagtccgTgtttgcAaatatcTcggactCaagccgcactgatt	325		
Db	644	aaatggaggagaaggaaagacaacgTatgcAAaacatcTaatgagagcAgaaaTcgatt	703		
QY	326	ctacacatgTgttgAACccgaggaaggaGcttatTgtaccagTatcatcgaggAccccaagc	385		

Db	704	ttattccaggagaaatcatgtgatatgtagactgtataattcacggatat---	760
Qy	386	aaatttaattccctcgatgtaccagggttaatttagagtatttaaatkattccctgttaga	445
Db	761	---tggtatttggatttccaagtgcgaagaggattttaaattatgggtatttaga	817
Qy	446	atgtgttaagaaacagggtcattcagttcggacaatacgaaccccttagttattcttgggtt	505
Db	818	ggaaatttaagaatggaaataatcgaacgggaaatacgaacccgttttgaagatag---	874
Qy	506	cgacacatgcatgtactcgtatttgcgaatgtcctgcctgattattgaaatacgcag	565
Db	875	agaaatgtcgaagtcattgttaatgggttaactttctccgaaggaagaattctttctga	934
Qy	566	ggacagaataaaactg	581
Db	935	agatcgaataaagtga	950

RESULT	l4
AAV24077	
ID	AAV24077 standard; DNA; lill BP.
XX	
AC	AAV24077;
XX	
DT	11-AUG-1998 (first entry)
XX	Banana bunchy top virus component 1
DE	
XX	
KW	BBTV; probe; diagnostic primer; com
XX	
OS	Banana bunchy top virus.
XX	
FH	Key
FT	Location/Qualifiers
FT	129..989
FT	/tag= a
XX	
PX	
PN	US5756708-A.
XX	
PD	26-MAY-1998.
XX	
PF	
PP	24-FEB-1994; 94US-0202186.
XX	
PR	24-FEB-1994; 94US-0202186.
XX	
PA	(YJQU-) UNIV QUEENSLAND TECHNOLOGY.

PI	Burns TW, Dale JL, Harding RM, Karan M;
XX	
XX	WPI; 1998-321636/28.
DR	P-PSDB; AAW54071.
DR	
XX	
XX	
PT	Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT	and primers and for producing virus-resistant plants
XX	
XX	Example 1; Fig 11; 59pp; English.
PS	
XX	
CC	This sequence represents a DNA fragment from the component 1 of
CC	the banana bunchy top virus (BBTV). The DNA molecules can be used
CC	as diagnostic probes or primers or can be inserted into plants or other
CC	organisms e.g. to produce virus-resistant plants or to act as promoters,
CC	enhancers or termination signals.
CC	
CC	
SQ	Sequence 1111 BP: 363 A; 192 C; 276 G; 280 T; 0 other;

```

Query Match          7.2%; Score 59.6; DB 19; Length 1111;
Best Local Similarity 51.3%; pred. No. 8.8e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;
QY 206 agtgcgtctgcgacgcgagagccacgcgcacatcctctggatgcgcagc 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy 206 agtgcgtctgcgacgagagccagcgcacacatccctctggtatgcggacg 265
 Db 584 agaggtggagaaatcatcgccagccatgtcatcgagagaaataatttgggtctatggccc 643
 Qy 266 agacggagagagcgggaagtcccggtgttccaaataatctcggactcaagcccgactgggt 325
 Db 644 aaatggagagagaggaagacaacgcatgcaaaacatctaatgaagacgagaaatgcgtt 703
 Qy 326 ctacacatgtgtgggaaccagaaagacgtattgtaccagtaacatcgagaccccaaacg 385
 Db 704 ttattctccaggaggaataatcattggatatgtagactgtataattacgagatat--- 760
 Qy 386 aaatttaactcgtatgcacgggtgtaatttagagattttaaattatgcctgttaga 445
 Db 761 ---tgttataattgatatccaaagatgcaagaggtattttaaattatgggttattaga 817
 Qy 446 atgtttaagaaacacgggcatcagttcggaacaaatacgaaccccttagttatcttgggtt 505
 Db 818 ggaatttaagaaatggaataatcacaagcgggaaatagaacccgttttgaagatag---t 874
 Qy 506 cgaccatgtgcgtactcgtatgtatttgcgaatgctcgtcctgattttgaaatacagcag 565
 Db 875 agaataatgcgaatgcattgttaattggctacttctccgaagaggaatcttttctga 934
 Qy 566 ggcagagaataaaactg 581
 Db 935 agatcgaataaagtgtg 950

Search completed: October 24, 2001, 10:07:16
 Job time: 621 sec

Db 611 agagtgagaaatcatcagcgcagccatgtcatcloggagaaataatttgggtctatggccc 670
 Qy 266 agacggagagagcgggaagtccgtgttgcgaataatctcggactcaagcccgactgggt 325
 Db 671 aaatggagagagaggaacacgtaagcaaacatcttaataagacgagaaatgcgtt 730
 Qy 326 ctacacatgtgtggaaccagaaagacgtattgtaccagtaacatcgagaccccaaacg 385
 Db 731 ttattctccaggaggaataatcattggatatgtagactgtataattacgagatat--- 787
 Qy 386 aaatttaactcgtatgcacgggtgtaatttagaattttaaattatgcctgttaga 445
 Db 788 ---tgttataattgatatccaaagatgcaagaggtattttaaattatgggttattaga 844
 Qy 446 atgtttaagaaacacgggcatcagttcggaacaaatacgaaccccttagttatcttgggtt 505
 Db 845 ggaatttaagaaatggaataatcacaagcgggaaatagaacccgttttgaagatag---t 901
 Qy 506 cgaccatgtgcgtactcgtatttgcgaatgctcgtcctgattttgaaatacagcag 565
 Db 902 agaataatgcgaatgcattgttaattggctacttctccgaagaggaatcttttctga 961
 Qy 566 ggcagagaataaaactg 581
 Db 962 agatcgaataaagtgtg 977

RESULT 15

AAV24084

ID AAV24084 standard; DNA; 1111 BP.

XX AC AAV24084;

XX DT 11-AUG-1998 (first entry)

XX DE Banana bunchy top virus component 1 DNA sequence.

XX KW BRTV; probe; diagnostic primer; component 1; ss.

XX OS Banana bunchy top virus.

XX PN US5756708-A.

XX PD 26-MAY-1998.

XX PF 24-FEB-1994; 94US-0202186.

XX PR 24-FEB-1994; 94US-0202186.

XX PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.

XX PI Burns TM, Dale JL, Harding RM, Karan M;

XX DR WPI; 1998-321636/28.

XX PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes

XX PT and primers and for producing virus-resistant plants

XX PS Disclosure; Column 17-18; 59pp; English.

XX CC This sequence represents a DNA fragment from the component 1 of

XX CC the Banana bunchy top virus (BRTV). The DNA molecules can be used as

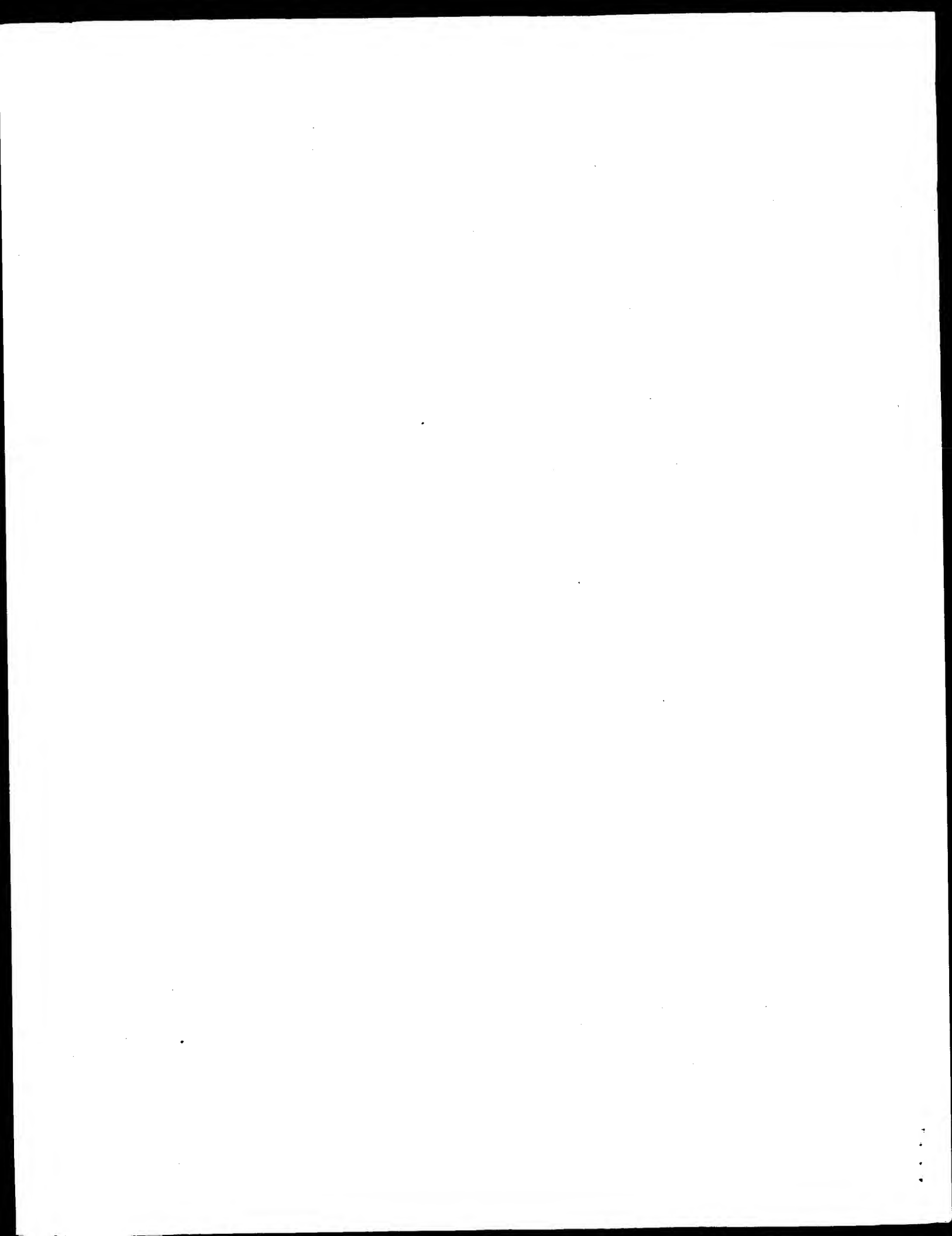
XX CC diagnostic probes or primers or can be inserted into plasmids or other

XX CC organisms e.g. to produce virus-resistant plants or to act as promoters,

XX CC enhancers or termination signals.

XX SQ Sequence 1111 BP; 363 A; 192 C; 276 G; 280 T; 0 other;

Query Match 7.2%; Score 59.6; DB 19; Length 1111;
 Best Local Similarity 51.3%; Pred. No. 8.8e-10;
 Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:40 ; Search time 5479.82 Seconds
(without alignments)
1436.950 Million cell updates/sec

Title: US-09-462-955-1_COPY_383_1215

Perfect score: 833

Sequence: 1 ggggtcttcgcagcagga.....accatccagtgagtagct 833

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

ppred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:594576"
/clone_lib="Stratagene ovarian cancer (#937219),"
/sex="female"
/dev_stage="adult, 64 years"
/lab_host="SOLR (kanamycin resistant)"
/notes="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Cloned unidirectionally. Primer: Oligo dt. Papillary
serous carcinoma, isolated from ascites, 64 year old
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGACTTTTTTTTTTTT 3'"
130 a 56 c 101 g 120 t 6 others
BASE COUNT
ORIGIN

```

	Query Match	4.68;	Score 38.2;	DB 3;	Length 413;	
	Best Local Similarity	54.7%;	Pred. NO.	0.59;		
	Matches	76;	Conservative	0;	Mismatches	63; Indels 0; Gaps 0;
Qy	101	ccagcgcataccgaagaatgcgttgtaacgagactcggtagaatgatacaagtctgaggc	160			
Db	140	CTGTGATGATCCGAAGAGAGAGTTCCTTTTAGGAGCTTCATTAAAGGCACAGAAGCGA	199			
Qy	161	cgtgaaattccgttcccatttcocatcataatggcagctttaagtctgctctgcgat	220			
Db	200	TTTTAAAAAGCCTAACCAAGGATTTCAGGTCCGGTTTGAGTCTTGAAGAATTTATTGCTTT	259			
Qy	221	cggagagccagcggacgat	239			
Db	260	TGAGCATCCTCGAAGAAGTT	278			

RESULT	3	
AA516896		
LOCUS	516 bp	mRNA
DEFINITION	vH97e09, l1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone IMAGE:902248 5', similar to SW:RS9_HUMAN P46781 40S RIBOSOMAL PROTEIN S9. ;, mRNA sequence.	
ACCESSION	AA516896	1
VERSION	AA516896.1	GI:22563355
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. l (bases 1 to 516)	
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	
TITLE	The WashU-HMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:52912 Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 473. Location/Qualifiers 1. .516 /organism="Mus musculus" /strain="C3H" /db_xref="taxon:10090" /clone="IMAGE:902248"	
FEATURES		
Source		

```

/clone_lib="Barstead mouse myotubes MRLRB5"
/lab_host="DH10B"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCAGATCGAATGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AATTCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."
BASE COUNT      117 a   126 c   154 g   118 t      1 others
ORIGIN

Query Match      4.6%; Score 38.2; DB 8; Length 516;
Best Local Similarity 51.8%; Pred. No. 0.62; 79; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 79;

QY      65 atttgctgaggaaacctgatgaactccgcttggagagaccaggcgatgataccgaagatcgct 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      38 ATTGTCAGCANACACATATGTGAACCCACCGGAGAACCCCTTCGGAGAAAGTTGGCGTCT 97

QY      125 tgtacaggagcttcggtggaatgagacaagatggcgcgctgaaataatccgttcccatctcc 184
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      98 TGGACCAAGAGCTAAAGTTGATTGGAGAGATGTGACATCCCGGAACCAACAGCTGAGGTTTGGGA 157

QY      185 atatcaaatggcgagcttgaagtgcgtctgcgatcgagagc 228
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      158 GGGTCAAGTTTACCTTGGCCCAAGATCCGTAAGCGCGCGGGGAGC 201

RESULT 4
AF0341173      2275 bp      mRNA      EST      30-MAR-1998
LOCUS      AF0341173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
DEFINITION      clone atccn2 contig. mRNA sequence.

```

```

RESULT      4
AF034173   2275 bp      mRNA      EST      30-MAR-1998
LOCUS      AF034173 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
DEFINITION clone ntcon2 contig, mRNA sequence.
ACCESSION  AF034173
VERSION    AF034173.1  GI:2707735
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 2275)
            Tripodis,N. and Ragoussis,J.
AUTHORS    Generation of a transcription map in the region immediately
TITLE       centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
            boundary
JOURNAL     Unpublished (1997)
COMMENT     Contact: Tripodis, Nikos
            Division of Medical and Molecular Genetics
            Guys Hospital
            7th floor, Guy's Tower, London SE1 9RT, UK
            Email: nikos@nki.nl.
FEATURES    Location/Qualifiers
             1..2275
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /map="6p21.3"
              /clone="ntcon2 contig"
              /clone.lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT  438 a 619 c 470 g 599 t 149 others
ORIGIN
Query Match 4.6%; Score 38; DB 14; Length 2275;
Best Local Similarity 16.7%; Pred. NO. 1.1;
Matches 32; Conservative 85; Mismatches 75; Indels 0; Gaps 0

```

QY	570	agaataaaactgtggaattattaaagattgtgtcatctaaattacacaaataccgcgcg	629
Db	1469	AAAAAAAAAAMWRYKRWKRKGRGRKRTGMYKRYRAMMAMCMMACWYWKRGMKKC	1528
QY	630	ccacgcgcctatgattacattatgaattctgcacagcgccgaaggcctggagggtg	689
Db	1529	WKYRKYKKYSTYKYSWSRYWTTTYYWCCTFSMKASASCAMRMWGYGSRSSRYW	1588
QY	690	ctaccgcgcgaagcgccggaacaatatgaatcggagttaggcggccgcacataaaag	749
Db	1589	GYGMSGCYGMTKRYRYSWTKWTWYMSMTHTWTTTWWRTTKTWWWW	1648
QY	750	attccattgga	761
Db	1649	WTTCTWTKRKGCA	1660
RESULT	5		
AQ615572			
LOCUS		498 bp DNA	
DEFINITION	HS 5144_B1.B02.SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=720 Col=3 Row=D, DNA sequence.	GSS	15-JUN-1999
ACCESSION	AQ615572		
VERSION	AQ615572.1	GI:5076848	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1. (bases 1 to 498) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.		
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome		
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)		
MEDLINE	99380589		
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 720 row: D column: 3 Seq primer: SP6 Class: BAC ends High quality sequence stop: 498. Location/Qualifiers 1. 498 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate-720 Col=3 Row=D" /clone_lib="RPCI-11 Human Male BAC Library" /sex="male" /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"		
BASE COUNT	171 a	107 c	87 g
ORIGIN		113 t	20 others
Query Match	4.4%	Score 36.8;	DB 231; Length 498;
Best Local Similarity	56.7%;	Pred. No. 1.7;	
Matches	68; Conservative	0; Mismatches	52; Indels 0; Gaps 0;

512	tggtgcatgactcgtatttggccaatgtcctgctgatttattgaaataatcagcaggagcaag	571
364	TATTCATGACATAGAAATTTGGGAAAGCATTTCCTCTATTATTAATTCAGTACGATCAT	305
572	aataaaactgtggaattattaaagtattgtcatctataaattacaccaa	619
304	GATATTTGCTGTGAGGTGGAATCATTTTATTTTAAATTAACACAA	257
RESULT 8		
AO089903/c	318 bp	DNA
LOCUS	HS_3001_A2_H04_MR C17	Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.
DEFINITION	AO089903	GSS
ACCESSION	AO089903.1	GI:3458814
VERSION		
KEYWORDS		
SOURCE	GSS.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 318) Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.	
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744	(1999)
MEDLINE	99380589	
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 3001 row: O column: 8 Class: BAC ends High quality sequence stop: 318.	
FEATURES		
source	Location/Qualifiers	
	1..318	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clones="Plate=3001 Col=8 Row=O"	
	/clone.lib="C17 Approved Human Genomic Sperm Library D"	
	/sex="male"	
	/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"	
BASE COUNT	77 a 66 c 66 g 108 t	1 others
ORIGIN		
Query Match	4.3%;	Score 35.6; DB 224; Length 318;
Best Local Similarity	54.6%;	Pred. No. 3.7;
Matches	71; Conservative	0; Mismatches 59; Indels 0; Gaps 0;
Qy	450 gtttaagaacaggcattcagtcaggcaataacgaaccccttgatttatcttgggtcgac	509
Db	212 GTTCTGCCAGGCGCAATCAGCGCAGGAGGAAGGATAATTAATTAGGAAGAGAG	153
Qy	510 catgtgcatgtactcgtatttgcgaatctcctgcgtgatttattgaaataatcagcaggac	569
Db	152 GATTTCAAATTTGTCCTGTTTGCAGTGACATGACTGTATATCTAGAAACCGCATTTGC	93
Qy	570 agaataaaac	579
Db	92 TCAGTCCAC	83
RESULT 9		
CNS016H0	11101 bp	DNA
LOCUS	CNS016H0	GSS
		26-JUL-1999

```

db      228    GTGTCAGAACTAACGTCTTACACCCTGCCCAATCA   263
||||| |||| | | | | | | | | | | | | | | | | | | | | |
RESULT  7
AZ444169          507 bp     DNA             GSS           04-OCT-2000
1M0239F20F mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0239F20 F, DNA sequence.
AZ444169
AZ444169.1 GI:10592981
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Musinae; Mus.
1 (bases 1 to 507)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D..Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: Q239 row: F column: 20
Seq primer: CGRTGTAATAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 507.
FEATURES                     Location/Qualifiers
     source                    .
         1..507
             /organism="Mus musculus"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="UUGC1M0239F20"
             /clone_lib="Mouse 10kb plasmid UUGC1M library"
             /sex="Male"
             /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
             /note="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gil4732114[gb]/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed into
adapted vector DNA, and transformed into
chemically-competent E.coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT                205 a       74 c       92 g       136 t
ORIGIN

```

Query Match 4.3%; Score 36; DB 244; Length 507;
Best Local Similarity 58.3%; Pred. No. 3.1; Indels 0; Gaps 0;
Matches 63; Conservative 0; Mismatches 45;

Thu Oct 25 13:08:06 2001

LOCUS CENS0000X 488 bp DNA GSS 28-JUN-1999
DEFINITION Arabidopsis thaliana genome survey sequence SP6 end of BAC F10F11
of IGF library from strain Columbia of Arabidopsis thaliana,
genomic survey sequence.
ACCESSION AL085215
VERSION AL085215.1 GI:5286355
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 488)
AUTHORS Salanoubat, M., Cholsne, N., Artiguenave, F., Brottier, P., Wincker, P.,
Samson, D., Saurin, W., Weissenbach, J. and Quetier, F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 488)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
FEATURES
source
Location/Qualifiers
1..488
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="IGF"
/clone="F10F11"
/note="end : SP6"
BASE COUNT 157 a 88 c 87 g 156 t
ORIGIN

Query Match 4.2%; Score 35.4; DB 219; Length 488;
Best Local Similarity 47.2%; Pred. No. 4.8;
Matches 108; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 447 tgggttaagaagcggcattcagtcgggacaaataacgacaccccttagtctctgggttc 506
|||||
Db 147 TCTTTTATCAACGACGGGTTTACCAGAGAGAGAGAGAAATACCTAGATCATCTTGATTC 206
|||||
QY 507 gaccatgtgcattctactgtattggcattgctgcctgattttgaaatacagcagg 566
|||||
Db 207 TTGTGTGACACTCTACATGATTGAATTTGTTTCCCAATTTACTTAACTTAAGAAATGAGT 266
|||||
QY 567 gacagaataaacgtggaattattaaagtattgtctatctctaaataacacaaataccgcg 626
|||||
Db 267 AATTGATGGAATTTAATTTGTTTGCAGGGTGTGTCACTAGTACAAATCTATGCAACC 326
|||||
QY 627 ccgccaagcgctatgctttacatcttataatctctgcccagggccga 675
|||||
Db 327 TCGCTCTGGTGCAATCATCCACTCTTGAACACATACCAAGGAGTCTGA 375
|||||

RESULT 12
BE603735 955 bp mRNA EST 02-MAR-2001
LOCUS BE603735
DEFINITION HVSMEH0087003f Hordeum vulgare 5-45 DAP spike EST library
HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0087003f,
mRNA sequence.
ACCESSION BE603735
VERSION BE603735.2 GI:13187835
KEYWORDS EST.
SOURCE EST.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 955)
AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and

Wood, T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
On Aug 21, 2000 this sequence version replaced gi:9861296.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCTCACTAAAGG
High quality sequence stop: 931.
FEATURES
source
Location/Qualifiers
1..955
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEH0087003f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library
HVCNDA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 303 a 205 c 213 g 220 t 14 others
ORIGIN

Query Match 4.2%; Score 35.4; DB 137; Length 955;
Best Local Similarity 52.0%; Pred. No. 5.7;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 268 acgagagagacgggaagtcctgtttgccaaatactctgagctcaagcccgaggttct 327
|||||
Db 585 ACGGGGACTGGGAAACCTCTGGGTTACCAACTAATCGGTTGACACATCCCTTTTC 644
|||||
QY 328 acacatgtgttggaaccgagaagcgtattgtaccaggtacatcgagagacccaaacgaa 387
|||||
Db 645 ACAGTTGGCGTATAACAAAAAGGCGGACGATTCCCTCCAAAAGTGCCCAATCTGAA 704
|||||
QY 388 atttaactctcgatgtaccaggtgaatt 417
|||||
Db 705 GGGGAATGGGAAATGAGACGCTAATATT 734
|||||

RESULT 13
AZ401897/c 345 bp DNA GSS 03-OCT-2000
LOCUS IM0168N22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0168N22 R, DNA sequence.
ACCESSION AZ401897
VERSION AZ401897.1 GI:10516971
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 345)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0168 row: N column: 22
 Seq primer: CACACAGAACACCTATGACC
 Class: plasmid ends
 High quality sequence stop: 345.
 Location/Qualifiers
 1. .345

FEATURES

source

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0168N22"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, P-"
 /notes="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114[gb]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
 104 a 77 c 58 g 106 t

BASE COUNT

ORIGIN

Query Match 4.2%; Score 35.2; DB 243; Length 345;
 Best Local Similarity 55.8%; Pred. No. 5;
 Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
 QY 416 tttagagattttaaattgcccctgttagaattgttgaagaacagggcatttcgga 475
 Db 330 TTTGAAGAGCTTAAATGGAGCGCTCGAATGAGTTTATAATGTTTACGTCATGA 271
 QY 476 caaatcgaaccccttagttatttgggttgaccattgtgcattgctattggcaa 535
 Db 270 AACCTGAGGACATCTTTTGTGGTGTAGTTTGACTGTCTACTTGTCTTGTGTGCA 211

RESULT 14

AV400883

LOCUS

DEFINITION

Bombyx mori brain Daizo P0 (just after pupation) Bombyx

mori cDNA clone br-1929 T3, mRNA sequence.

ACCESSION

AV400883

VERSION

EST.

KEYWORDS

domestic silkworm.

SOURCE

Bombyx mori

ORGANISM

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia

; Bombycoidea; Bombycidae; Bombyx.

REFERENCE

1 (bases 1 to 793)

Mita, K., Morimoto, M., Shimada, T., Okano, K. and Maeda, S.

Bombyx mori cDNA

Unpublished (2000)

JOURNAL

Contact: Mita K

Genome Research Group

National Institute of Radiological Sciences

Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

Email: knita@nirs.go.jp

method: uni-directional, sequence direction: sequenced from T3 primer
 (5' -> 3')

Project=Silkworm Genome Program in MAFF, and Research for the
 Future Program in JSPS. see 'SilkBase',
 <http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.

FEATURES

source

1. .793
 /organism="Bombyx mori"
 /strain="Daizo"
 /db_xref="taxon:7091"
 /clone="br-1929"
 /clone_lib="Bombyx mori brain Daizo P0 (just after pupation)"
 /sex="female/male mixed"
 /tissue_type="brain"
 /dev_stage="P0 (just after pupation)"
 198 a 231 c 223 g 141 t

BASE COUNT

ORIGIN

Query Match 4.2%; Score 35.2; DB 30; Length 793;
 Best Local Similarity 51.2%; Pred. No. 6.3;
 Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
 QY 167 aaatccgttccatttccatcaccaattggcagcttgaagtgtctgtcgatcgaga 226
 Db 634 AAACCTGCACCAACCAAGCAGCGTCTCCAGTCCGAGGTCGGCCAGACGCTGCA 693
 QY 227 gccacgagcgcacacatctctggatctgcagacgagcagcagcagcagcagc 286
 Db 694 CCACACAGAGAGCTGTTAAGCCAGAGAAATAGTAGAGATCGAGAGAGAGAG 753
 QY 287 cgtgttgccaaatctcgcgactcaagccgactggttc 326
 Db 754 TTAGATGGCAGCGTCCAGATGAAGGTCCGCCCGGTC 793

RESULT 15

A0746046

LOCUS

DEFINITION

HS_2277_A2_F11_T7C CIT Approved Human Genomic Sperm Library D Homo

sapiens genomic clone plate=2277 Col=22 Row=K, DNA sequence.

ACCESSION

A0746046

VERSION

GSS

KEYWORDS

human.

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 848)

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and

Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

CONTACT: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: http://www.htsc.washington.edu

Plate: 2277 row: K column: 22

Seq primer: T7

Class: BAC ends

High quality sequence stop: 848.

Location/Qualifiers

1. .848

/organism="Homo sapiens"

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:16 ; Search time 180.6 Seconds
(without alignments)
873.179 Million cell updates/sec

Title: US-09-462-955-1_COPY_383_1215
Perfect score: 833
Sequence: 1 gggtgtcttcgagcagcga.....accatccagtgagtagct 833

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	120.6	14.5	1096	2	US-08-418-071-4
2	117	14.0	1091	2	US-08-418-071-5
3	102.2	12.3	1106	2	US-08-418-071-3
4	99.2	11.9	1017	4	US-08-793-634B-6
5	91.2	10.9	1022	4	US-08-793-634B-2
6	62.8	7.5	1110	1	US-08-202-186-14
7	59.6	7.2	982	3	US-08-973-068-28
8	59.6	7.2	1110	1	US-08-202-186-11
9	59.6	7.2	1111	1	US-08-202-186-9
10	59.6	7.2	1111	1	US-08-202-186-12
11	59.6	7.2	1111	1	US-08-202-186-24
12	58.6	7.0	1103	1	US-08-202-186-16
13	58.6	7.0	1105	1	US-08-202-186-18
14	58	7.0	1109	1	US-08-202-186-13
15	58	7.0	1111	1	US-08-202-186-15
16	57	6.8	1104	1	US-08-202-186-17
17	56.4	6.8	1111	1	US-08-202-186-10
18	46.4	5.6	287	2	US-08-418-071-1
19	46.4	5.6	300	2	US-08-418-071-2
20	32.2	3.9	7218	1	US-08-232-463-14
21	31	3.7	1212	3	US-09-046-578-5
22	30.4	3.6	1212	3	US-09-046-578-1
23	29.4	3.5	859	3	US-09-280-409-2
24	29.4	3.5	1414	4	US-08-686-968C-228
25	29.4	3.5	1437	3	US-08-724-814-15
26	29.4	3.5	8920	2	US-08-446-855A-1
27	29.4	3.5	8920	4	US-09-150-741-1

c	28	29.2	3.5	1209	6	5352575-4	Patent No. 5352575
	29	29	3.5	9468	1	US-08-325-547-10	Sequence 10, Appl
	30	28.8	3.5	1596	5	PCT-US94-11328A-3	Sequence 3, Appl
c	31	28.8	3.5	10607	1	US-08-078-090-3	Sequence 3, Appl
c	32	28.6	3.4	1593	2	US-08-524-828-2	Sequence 2, Appl
c	33	28.6	3.4	1593	2	US-08-524-828-2	Sequence 2, Appl
c	34	28.6	3.4	1593	2	US-08-975-114A-2	Sequence 2, Appl
c	35	28.6	3.4	1608	2	US-08-849-281A-2	Sequence 2, Appl
c	36	28.6	3.4	1608	2	US-08-524-828-3	Sequence 3, Appl
c	37	28.6	3.4	1608	2	US-08-975-114A-3	Sequence 3, Appl
c	38	28.6	3.4	1608	2	US-08-849-281A-1	Sequence 1, Appl
c	39	28.6	3.4	2247	2	US-08-524-828-1	Sequence 1, Appl
c	40	28.6	3.4	2247	2	US-08-975-114A-1	Sequence 1, Appl
	41	28.6	3.4	2899	2	US-08-624-581-2	Sequence 2, Appl
	42	28.6	3.4	2917	2	US-08-624-581-3	Sequence 3, Appl
	43	28.6	3.4	2959	2	US-08-624-581-1	Sequence 1, Appl
	44	28.4	3.4	1380	4	US-09-499-505-4	Sequence 4, Appl
	45	28.4	3.4	1430	4	US-09-499-505-3	Sequence 3, Appl
c	45	28.2	3.4	1556	2	US-08-881-857-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rev-Yuh
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418.071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-4

Query Match 14.5%; Score 120.6; DB 2; Length 1096;
Best Local Similarity 52.6%; Pred.No. 4.1e-30;
Matches 317; Conservative 0; Mismatches 274; Indels 12; Gaps 2;

y y 7 tctcgcagccagcgactccgactcgtcctggagtccaaggcccacgattggcgcacaacgat 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 348 TTCCTGAACCTGGGTACTCCGTTGGTTCCTGGTTCGAAGAAGCGCAAGCTTCCTCATAGAT 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 67 ttgtcggagaacctgatgaacctccgcctcggtaagaccgccagcgatataccgaagatgcttg 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 408 TCAGACAGAGCCCTCAGGAATTTGAAGATGGACGATGCCAAGTATCGCAGATGCTTG 467
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 127 tacacggagcttcgtygaaatgagacaaatagggccgctgaaaattccgttccccattcccat 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 468 CAGTGGAAATCAAAATAAAGATGCCAGAAATTAATTCGGAATGGGTTACGAACTA ----- 520
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 187 atcaaatcggcagcttggaagtgcctgctgcgatalcggagagccagcgagcgtcgcaaa 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 521 --AAAGANTGGCAAAATAAATAATTCAACACATCGAAGGTGTTCTGTGATGCAGAGTA 578
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 247 tcctctggatatcggcacgagacgagcgagagcgaggaagtcogtgtttgccaataatctcg 306
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 579 TCATCTGGGTATACGGTCCCAACGGAGCGAGGAAGTCAACCTTCGCAAGATATCTAT 638
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 307 gactcaagcccgactggtttcacatgctggtggaaaccagaaaggacgtattgtaccagt 366
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 639 CATTAAMACC GGATCGGGATATATCAACGGTGGAAACGCTCGGATATGATGCACATCA 698
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 367 acatcggagaccacaaacgaatttaactcctcgtatgccacagtgtaatttagatt 426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 699 TAACGATGGATCCTGATAATCATTTGGATATTGTATATCCCAGAAGTCATTCAGATTATC 758
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 427 taaattatgcctgttagaatgtgttaaagacagggcattcagttcggacaaatcacgaac 486
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 759 TGAATTTATGGCGTTATAGAACAAATTAACGANATAGAGTTTTTANTAATAACAAAATACGAAAC 818
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 487 c---ccttagttatcttgggttcgaccatgtgcattgactcgtatttgcccaatgctcgc 543
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 819 CATGTGTCATTAGAAAGATGGACAAAATGTCATGTAATTTATGGCAAATGTTGTC 878
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 544 ctgattattgaaaaatcagcagcgagacagaataaaactgtggaattttaagatgltgc 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 879 CTGATTATTCTAAAAATTTTCAGAGATAGATAAAAATAAATTAATTGTTTCAGAAAGGAAC 938
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y y 604 atc 606
||
b 939 TTC 941

RESULT 2
US-08-418-071-5
; Sequence 5, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0

Query Match	12.3%;	Score 102.2;	DB 2;	Length 1106;
Best Local Similarity	53.3%;	Pred. No. 5e-24;		
Matches 215;	Conservative 0;	Mismatches 188;	Indels 0;	Gaps 0;
195 tggcagcttgaagtgtgtctgtcgtatcgagagccagcggacgatgcacaaatcctctgg	254			
524 TGCAGAGATTCAAATTGACGGAGCAATTGACGAGAACCCGATGATCGAAGCATCATCTGG	583			
255 atatgcgcagcagacagggagagcaggggaagtcocgtgtttgccaaatatctcggactcaag	314			
584 GTCTATGGTCCTTATGGTAAATGAGGTAATCAACATATGCGAAGTCACTAATCAAGAAG	643			
315 ccgcgactgggtctacacatgtgtgtggaaccagaaaggacgtattgtaccagtcacatcag	374			
644 GATTGGTTCTACACGAGGGTGGGAAGAGAGATATCTTATTCCTATCTGGAGCAA	703			
375 gaoccaaaacgaatttaalccctcgtatgcaccagcgggtgaatttagagtatttaaatat	434			
704 GGATCTGCACAAGCATATAGTATTTCATATTCCTCGTGTGAATCAGGATTATTTAAATAT	763			
435 gccctgttggaatgtgtttaagacacaggggcattcagttcggacaaatacgaacccttaqt	494			

QY 309 ctcaagcccgactgtttctacacatgtgttggaaccagaaagacgtattgtaccagttac 368
 Db 618 TTGAAGATGTTGGGTATCTCCCTGGAGGAAAGACACAAGATATGATGCATCTTGTG 677
 QY 369 atcagagagcccaaaacaaatttaactcctcgatgtaccagaggtgttaatttagagtattta 428
 Db 678 ACTGCTGAGCCCTAGAAATAAATGGGTATTTGACATACCCAGAGTAGTTCCAGAGTAGTGTG 737
 QY 429 aattatccctgttagaattgtttaagaacaggcgattccagttcgggacaaatacgaacc- 487
 Db 738 AATTAAGTGTAAAGACACAGCTTAAGAAAGAGTAAATGTTGTAATCACTAGTATGAGCCA 797
 QY 488 --ccttagttatcttgggttcgaaccatgtgcattcgtatcgtattgtccaaatgtcctgcct 545
 Db 798 TGTGTAATCGGGATGATAATCATCTCTTTCATGTAATTTGTTGCAAAATGTAATCCCA 857
 QY 546 gattattgaaatacagcagagagcagaaataaaactgtggaatttttaaaagtattgt 601
 Db 858 GATTTGGGAAAATAAAGTGAAGTAGAATAAAATAATTCGTTGTTGTAATAACTCTG 913

RESULT 5

US-08-793-634B-2
 ; Sequence 2, Application US/08793634B
 ; Patent No. 6211431
 ; GENERAL INFORMATION:
 ; APPLICANT: Boevink, Petra C.
 ; APPLICANT: Surin, Brian P.
 ; APPLICANT: Keese, Paul K.
 ; APPLICANT: Chu, Paul W.G.
 ; APPLICANT: Waterhouse, Peter M.
 ; APPLICANT: Khan, Rafiqul I.
 ; APPLICANT: Larkin, Philip J.
 ; APPLICANT: Taylor, William C.
 ; APPLICANT: Marshall, Jerry S.
 ; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/793,634B
 ; FILING DATE: June 9, 1997
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digiglio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 10530
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 516-742-4343
 ; TELEFAX: 516-742-4366
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1022 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-793-634B-2

Query Match 10.9% Score 91.2; DB 4; Length 1022;
 Best Local Similarity 53.5%; Pred. No. 2,1e-20;
 Matches 214; Conservative 0; Mismatches 183; Indels 3; Gaps 1;

QY 195 tggagcgttgaagtgtctgtcgatcgagagccagcgagcgtatcgacacatcctctgg 254
 Db 529 TGGCAAAATTTAGCTTTCACAGGATTTAATGGCGGAACACAGATGATCGGAGTATCATCTGG 588
 QY 255 atatgcgagcagagcgagcgagcgagtcctgtgtttggccaaatactcggagactcaag 314
 Db 589 GTCTATGTTTCAGACGGGAGCAAGAAAGACGAGCTTCGCCGAAGGAATT--AATCAGG 645
 QY 315 ccgcactgtttacacatgtgtgtggaaccagaaagacgtattgtaccagttacatcgag 374
 Db 646 TATGGATGCTTTTATACAGCCGGAGGAAAGACCCAGAGCGTATTATATATGATGCTCAA 705
 QY 375 gaccacaaagaaatttaactcctcgaatgtaccaggtgttaatttagatttaattat 434
 Db 706 GACCCAGAGAGGAGTAATTCGGTTTGAIGTTCACAGGTGTTCTTCGGAGATGATGAACAT 765
 QY 435 gcctgttagaattgtttaagaacaggcgattcagttcgggacaaatacgaacccctagt 494
 Db 766 CAGCGCATGGAGATGTTGAAGAACAGAGTTTTCGAAGTACAAAATATAGGCTGTAGAT 825
 QY 495 tatcttgggttcgacctgtgcattcgtatcgtattgtccaaatgtcctgcctgattttg 554
 Db 826 CTTTGTATTAGGAAGTATGTTCAATTTAATTTGTTTGGCCAACTGGCACCTGACCCACG 885
 QY 555 aaatcgcagcgagcagacagataaaactgtgtgaattatttaa 594
 Db 886 CGCATAGTGGAGACAGACTTGTAAATATCAATTTGTTGAA 925

RESULT 6

US-08-202-186-14
 ; Sequence 14, Application US/08202186
 ; Patent No. 5756708
 ; GENERAL INFORMATION:
 ; APPLICANT: KARAN, Miro
 ; APPLICANT: BURNS, Thomas M.
 ; APPLICANT: DALE, James L.
 ; APPLICANT: HARDING, Robert M.
 ; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
 ; NUMBER OF SEQUENCES: 60
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W.
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/202,186
 ; FILING DATE: 24-FEB-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: JEFFERY, Donald D.
 ; REGISTRATION NUMBER: 19,980
 ; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202 672 5300
 ; TELEFAX: 202 672 5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1110 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; US-08-202-186-14

Db 230 ---tggtatatattccaagatgcgaagagattatttaattatgggtattaga 286
QY 446 atgtgttaagaacaggcatttcgtagcaaatcaccaaccttagttattcttgggt 505
Db 287 ggaatttaagaattggaataattcaaacgcggaataatgaaccttctgaagatag ---t 343
QY 506 cgacctatgcctactcgtatttgcgaatgcctcgcctgattattgaaatcagcag 565
Db 344 agaatagtcgaagtcatttgtatggcgaactctctcgaaggaaggaattcttctga 403
QY 566 ggcagaataaaaactg 581
Db 404 agatcgaataaaattg 419

```

RESULT      8
US-08-202-186-11
: Sequence 11, Application US/08202186
: Patent No. 5756708
: GENERAL INFORMATION:
: APPLICANT: KARAN, Mirko
: APPLICANT: BURNS, Thomas M.
: APPLICANT: DALE, James L.
: APPLICANT: HARDING, Robert M.
: TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
: NUMBER OF SEQUENCES: 60

```

ADDRESSEE: Forey & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/202,186

```

STRANDEDNESS: Single
TOPOLOGY: circular
S-08-202-186-11

Query Match          7.2%; Score 59.6; DB 1; Length 1110;
Best Local Similarity 51.3%; Pred. No. 6.1e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps

Y 206 agtctgtctgcgacgcggagagccagcgacgcacgaatcctctggatatcgcgacg 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 584 AGAGTGTGAGAAATATCATGCGCAGCCATGTCTCGAGAAATAATTTGGGTCTATGCACC 643
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 266 agacggaggagacggdaagtcgcggtgttgccaaatctctcggactcaagccgcgtggt 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 644 AAATGGAGGAGAGGAAGACAACCATCTATCCAAACATCTAATGAGACGAGAAATCGGTT 703
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 326 ctacacatgttgatggagaccagaaagacgctattgtaccagtcacatcgaggaccctaaacg 385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 704 TTATCTCCAGGAGGAAATATCATGGATATGTAGACTGTATAATATCAGAGGATAT--- 760

```

	Query Match	7.2%	Score 59.6;	DB 1:	Length 11111;
	Best Local Similarity	52.6%;	Pred.	No. 6.1e-10;	
	Matches 181;	Conservative	0;	Mismatches 154;	Indels
					Gaps
QY	238	atcgcacaaatctctggatatcgagacgagacgagagagaggaagtccgtttaccaca	297		
Db	616	ATTCGGAGATAAATTTGGGTCTATGCCCCAAATCGAGGAGGAAGARACACACGTATGCCAA	675		
QV	298	aatactctcgactcaagcccgaactggttctcacacatgtggttggaaccagaaggsgtgtat	357		

[illegible]

Db 676 AACAACTAATGAGCAGGAGGATGCTTTTATCTCCAGGGGAAATCAITGGATATAT 735
QY 358 tgtaccagtcacatcgagagaccacaaatatttaactcctcgatgacccaggtgtaatt 417
Db 736 GTAGACTGTATATTAACGAGGATAT-----TGTATATTTGATATTCGAAGATGCAAG 789
QY 418 tagagtatttaaatatgacctggttagaatgtgttaagaacagggcattccagttcggaca 477
Db 790 AGGATATTTAAATTAATGCTTATTAGAAGAAATTAAGAATGGAATAATTCAAACCGGGA 849
QY 478 aatcgaacaccttagttatcttgggttcgaccatgtgcacatgcatgtactogtatttcccaatg 537
Db 850 AATATGAACCGCTTTTGAAGATAG---TAGAATATGTCGAAGTCAATGTAATGCTAACT 906
QY 538 tcttcctgattatttgaataatcgagcaggggacagaaataaaaaactg 581
Db 907 TCCTTCCGAGGAGGAATCTTTTCTGAGATCGAATAAAGTTG 950

RESULT 11

US-08-202-186-24
; Sequence 24, Application US/08202186
; Patent No. 5756708

GENERAL INFORMATION:

; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA

; ZIP: 20007-5109

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994

; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FTKE

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202 672 5300

; TELEFAX: 202 672 5399

; TELEX: 904136

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

; LENGTH: 1111 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: circular

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 129...989

US-08-202-186-24

Query Match

Best Local Similarity 51.3%; Score 59.6; DB 1; Length 1111;

Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 206 agtctgtctcgatcgagagccagcgacatcgacaaatcctctgatatcgagcagc 265

Db 611 AGAGGTGGAATATCATCGGCGACCCATGTCTATCGGAGAAATAATTTGGGTCTATGCCCC 670

QY 266 agacggaggagaggggagccgtgttgcacaaatattctcgactcaagcccgactggtt 325
Db 671 AAATGGAGGAGAGGAAAGACAACAGTATGCAAAACATCTAATGAAGCAGAGAAATCGGTT 730
QY 326 ctacacatgtgttggaaaccagagaagagcgtattgtaccagtlacatcgagggaccccaaacg 385
Db 731 TTATTCFCCAGGAGGAAATCAATGGATATATGTAGACTGTATAATACGAGGATAT--- 787
QY 386 aaatttaactcctgtaccaggtgttaattttagagatttttaataattgcccctgttaga 445
Db 788 ---TGTTATATTTGATATTCAGATGCAAGATGCAAGAGGATATTTAAATATGGGTATTAGA 844
QY 446 atgtgttaagaacagggcattcagttcggacaaataacgaaccccttagttatcttgggtt 505
Db 845 GGAATTTAAGAATGGAATAATTCAAAGCGGGAATATGAACCGCTTTTGAAGATAG---T 901
QY 506 cgaccatgtcagtcactcgtatttgcacaaatgtcctgcctgattattgaaaaatcagcag 565
Db 902 AGAATATGTCGAAGTCAITGTATGCTAACTTCCTTCCGAAGGAGGAATCTTTTCTGA 961
QY 566 ggacagaaataaaaaactg 581
Db 962 AGATCGAATAAAGTTG 977

RESULT 12

US-08-202-186-16

; Sequence 16, Application US/08202186

; Patent No. 5756708

GENERAL INFORMATION:

; APPLICANT: KARAN, Mirko

; APPLICANT: BURNS, Thomas M.

; APPLICANT: DALE, James L.

; APPLICANT: HARDING, Robert M.

; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W.

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/202,186

; FILING DATE: 24-FEB-1994

; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: JEFFERY, Donald D.

; REGISTRATION NUMBER: 19,980

; REFERENCE/DOCKET NUMBER: 71611/102 FTKE

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202 672 5300

; TELEFAX: 202 672 5399

; TELEX: 904136

INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1103 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: circular

US-08-202-186-16

Query Match

Best Local Similarity 52.5%; Score 58.6; DB 1; Length 1103;

Matches 180; Conservative 0; Mismatches 154; Indels 9; Gaps 2;

Db 584 AGAGGTGGAGAAATATCATGCGCGCAGCCATGTCATCGGAGAAATAATTTGGGTCTATGGCCC 643
 QY 266 agacggagacacgggaagtcggtgttgccaaatctcgcagctcagcccgactggtt 325
 Db 644 AAATGGAGGAGAAAGAAAGCAACGTATGCAAAACGCTTAATGAAGACGAGAAATGGGTT 703
 QY 326 ctacacatgtgtggaacgaggaagcgtattgtaccagttacatcgagcccaaacg 385
 Db 704 TTATTTCCAGGAGGAAATCATTTGGATATATGTAGACTGTATAATTACGAGGATAT--- 760
 QY 386 aaatttaactcgcagtcaccaggtgtaatttagagttattaaattatgccctgttaga 445
 Db 761 ---TGTTATATTTGATATTTCAAGATGCAAAAGAGGATTAATTAATTTATGGTTATTAGA 817
 QY 446 atgtgttaagaacagggcattcagtcggacaaatacgaaccccttagttatcttgggtt 505
 Db 818 GGAATTTAAGAAATGGAATAATTCAAACGGGAAATATGAACCCGTTTGAAGATAG---T 874
 QY 506 cgaccatgtgcagtcactgtatttgcgaatgtcctgcctgattttgaaatacagcag 565
 Db 875 AGAATATGTCGAAGTCATTTGTAATGCTAATTCCTTCCGAGGAAGGAATCTTTCTGA 934
 QY 566 ggacagaataaaactg 581
 Db 935 AGATCGAATAAAGTTG 950

RESULT 15
 US-08-202-186-15
 ; Sequence 15, Application US/08202186
 ; Patent No. 5756708
 ; GENERAL INFORMATION:
 ; APPLICANT: KARAN, Mirko
 ; APPLICANT: BURNS, Thomas M.
 ; APPLICANT: DALE, James L.
 ; APPLICANT: HARDING, Robert M.
 ; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
 ; NUMBER OF SEQUENCES: 60
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W.
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/202,186
 ; FILING DATE: 24-FEB-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: JEFFERY, Donald D.
 ; REGISTRATION NUMBER: 19,980
 ; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202 672 5300
 ; TELEFAX: 202 672 5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1111 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; US-08-202-186-15

Query Match 7.0%; Score 58; DB 1; Length 1111;
 Best Local Similarity 51.1%; Pred. No. 2,1e-09;
 Matches 192; Conservative 0; Mismatches 175; Indels 9; Gaps 2;

QY 206 agtctgtctgcgcatcgcgagccagcgacgacatcgacaaatcctctggtatgacgagc 265
 Db 584 AGAGGTGGAGAAATATCATGCGCGCAGCCATGTCATCGGAGAAATAATTTGGGTCTATGGCCC 643
 QY 266 agacggagacacgggaagtcggtgttgccaaatctcgcagctcagcccgactggtt 325
 Db 644 AAATGGAGGAGAAAGAAAGCAACGTATGCAAAACGCTTAATGAAGACGAGAAATGGGTT 703
 QY 326 ctacacatgtgtggaacgaggaagcgtattgtaccagttacatcgagcccaaacg 385
 Db 704 TTATTTCCAGGAGGAAATCATTTGGATATATGTAGACTGTATAATTACGAGGATAT--- 760
 QY 386 aaatttaactcgcagtcaccaggtgtaatttagagttattaaattatgccctgttaga 445
 Db 761 ---TGTTATATTTGATATTTCAAGATGCAAAAGAGGATTAATTAATTTATGGTTATTAGA 817
 QY 446 atgtgttaagaacagggcattcagtcggacaaatacgaaccccttagttatcttgggtt 505
 Db 818 GGAATTTAAGAAATGGAATAATTCAAACGGGAAATATGAACCCGTTTGAAGATAG---T 874
 QY 506 cgaccatgtgcagtcactgtatttgcgaatgtcctgcctgattttgaaatacagcag 565
 Db 875 AGAATATGTCGAAGTCATTTGTAATGCTAATTCCTTCCGAGGAAGGAATCTTTCTGA 934
 QY 566 ggacagaataaaactg 581
 Db 935 AGATCGAATAAAGTTG 950

Search completed: October 24, 2001, 10:00:19
 Job time: 204 sec

us-09-462-955-1_copy_383_1215.rni

Thu Oct 25 13:08:04 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:43 ; Search time 5479.82 Seconds
(without alignments)
1347.248 Million cell updates/sec

Title: US-09-462-955-1_COPY_211_991
Perfect score: 781
Sequence: 1 gttgggacgaggttgaccc.....ttaagatgtgtcatctaa 781

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

/clone_lib="Barstead mouse myotubes MPLRB5"
 /cell_line="C2C12"
 /lab_host="DH10B"
 /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTTCAGATCTGAAGTGGAGCGCCGCTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 [AATTCGATCCCTTG], digested with Not I and cloned into the
 Not I and Eco RI sites of the modified pT7T3 vector.
 Library constructed by Bob Barstead. The C2C12 cell line
 (available from ATCC, catalog # CRU-1772) differentiates
 rapidly, forming contractile myotubes and producing
 characteristic muscle proteins."

BASE COUNT	ORIGIN
117 a	126 c
154 g	118 t
1 others	

	Query Match	4.9%	Score 38.2	DB 8	Length 516
	Best Local Similarity	51.8%	Pred. No. 0.62		
	Matches 85	Conservative 0	Mismatches 79	Indels 0	Gaps 0
Qy	237	atttgcctgaggaaacctgatgaactccgcctggaagaccgagcggtaccgaagaatgcgt	296		
Db	38	ATTGTGACANAACTATGTAAACCCACGGGAGAACCCCTTCGGAGAAAGTTGGCGTCT	97		
Qy	297	tgtacacgagccttcgctggaaatggacaagaatggcgctgaaatccgcttccccattcc	356		
Db	98	TGGACACGAGCTAAAGTTGATTGGAGATATGGACTCCGGAACAAACGTCAGGTTTGA	157		
Qy	357	atatcacaaattggcagccttgaaagtctctcgacacgagagc	400		
Db	158	GGGTCAAGTTTACCCCTGGCCAGATCCCTGAAGCGCCCGGGAGC	201		

RESULT	4
AQ615572	
LOCUS	498 bp DNA
DEFINITION	HS_5144_B1.B02.SP6E.RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=720 Col=3 Row=D, DNA sequence.
ACCESSION	AQ615572
VERSION	AQ615572.1
KEYWORDS	GI:5076848
SOURCE	GSS.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 498)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., an Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

<http://www.nslc.washington.edu>
 Plate: 720 row: D column: 3
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 498.
 Location/Qualifiers

```

source
1. .498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=720 Col=3 Row=D"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT      171 a      107 c      87 g      113 t      20 others
ORIGIN

Query Match      4.7%; Score 36.8; DB 231; Length 498;
Best Local Similarity 56.7%; Pred. No. 1.7;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 622 gtaagaacaggcattcagtcgacaaataacgaccccttagtattcttggttcgac 681
||| | ||||| |||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 271 GTCTGGCCAGGCAATCAGGCAAGCAAGCAAAATAAGGGTATTAACTTAGGAAAAGAG 330
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 682 catgtgatgtactgtatttgcacaaatgcctgctgatttgaataatcagcaggac 741
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 GAAGTCAAAATGTCCTGTTGTCAGATGACATGATGTTTATTATTAGAAATCCCATCGAC 390
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
AW818171      518 bp      mRNA      EST      17-MAY-2000
LOCUS      CMI-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.
ACCESSION      AW818171
VERSION      AW818171.1 GI:7911165
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 518)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4t2-CMI-ST0277-161
299-070-h05&t3=1999-12-16&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 76
High quality sequence stop: 183.
Location/Qualifiers
1. .518
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0277"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
source

FEATURES
source
1. .318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3001 Col=8 Row=O"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      77 a      66 c      66 g      108 t      1 others
ORIGIN

Query Match      4.6%; Score 35.6; DB 224; Length 318;
Best Local Similarity 54.6%; Pred. No. 3.6;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 622 gtaagaacaggcattcagtcgacaaataacgaccccttagtattcttggttcgac 681
||| | ||||| |||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 212 GTTCTGCCAGGCAATCAGGCAAGCAAGCAAAATAAGGGTATTAACTTAGGAAAAGAG 153
||| | ||||| |||| | ||||| | ||||| | ||||| | ||||| | |||||

```

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 193 a 119 c 89 g 117 t

ORIGIN

Query Match 4.7%; Score 36.8; DB 121; Length 518; Best Local Similarity 61.5%; Pred. No. 1.7; Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 545 aggaccccaaacgaaatttaactcgtacgtaccagggtgaatttagatttaattaaatt 604
|||| ||||| ||||| |||| ||||| ||||| |||| ||||| ||||| |||||

Db 168 AGCATTAAACCAATTTTCATCTCACATGGACACCAAGCAAAAATATATAGTACAAAGATT 227
|||| ||||| ||||| |||| ||||| ||||| |||| ||||| ||||| |||||

Qy 605 atgcctgttagaatgtgttaagaacaggggcattca 640
|||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 228 GTCTCAGAGTAACGCTCTTAACACCTGCCCAATCA 263
|||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
AQ089903 318 bp DNA GSS 26-AUG-1998
LOCUS HS_3001_A2_H04_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.
ACCESSION AQ089903
VERSION AQ089903.1 GI:3458814
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 318)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3001 row: O column: 8
Class: BAC ends
High quality sequence stop: 318.
Location/Qualifiers
1. .318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3001 Col=8 Row=O"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 77 a 66 c 66 g 108 t 1 others

ORIGIN

Query Match 4.6%; Score 35.6; DB 224; Length 318; Best Local Similarity 54.6%; Pred. No. 3.6; Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 622 gtaagaacaggcattcagtcgacaaataacgaccccttagtattcttggttcgac 681
|||| ||||| ||||| |||| ||||| ||||| |||| ||||| ||||| |||||

Db 212 GTTCTGCCAGGCAATCAGGCAAGCAAGCAAAATAAGGGTATTAACTTAGGAAAAGAG 153
|||| ||||| ||||| |||| ||||| ||||| |||| ||||| ||||| |||||

were sequenced additional times to assess quality

```

QY 682 catgtgcatgtactgtatttgcacaaatgctgctgctgatttatttgaataatcagcaggac 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 GATTTCAAAATGTCCTGTTTGCAGATGACATGACTGTATATCTAGAAAACCGCATGTC 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 742 agataaaac 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92 TCAGTCCAAC 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
A1942816 460 bp mRNA EST 06-AUG-1999
LOCUS f657b05.xl zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA
DEFINITION sequence.
ACCESSION A1942816
VERSION A1942816.1 GI:5707472
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 460)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,F., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
WashU zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourcencentrumPrimarbatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T7 ET from Amersham
High quality sequence stop: 443.
FEATURES
source
1. 460
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'-pGACTAGTCTAGATCCGAGCGCGCCCTTTTCTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones

```

BASE COUNT 154 a 60 c 59 g 187 t

ORIGIN

Query Match 4.5%; Score 35.4; DB 104; Length 460;
 Best Local Similarity 50.9%; Pred. No. 4.6;
 Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 595 tatttaattatgcctctgttagaattgtttaagacagcaggcatttcgaatgcgacaaatc 654
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 5 TTTTITTTTTTACAAATTTGAAGTGTGTTTACAAAAACAATTAATTTGTCCAAAAA 64
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 QY 655 gaaccccttagttatctctgggttcgaccatgtgcattgtactctgatttgcgaatgcctg 714
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 65 AACCTCCATAATATATGTTGTTTAAAGCTCTATTATAATTCAGTAGTTTGAACAAACACCAAG 124
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 QY 715 cctgattatttgaaatcagcaggcagcagataaaactgtggaat 759
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 125 TATAATTTTGTGAATTCATCTAGTGAATATATATTTCATGTCGCAT 169
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8

BE603735 955 bp mRNA EST 02-MAR-2001

LOCUS HVSMEH0087003f Hordeum vulgare 5-45 DAP spike EST library

DEFINITION HVSMEH0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0087003f, mRNA sequence.

ACCESSION BE603735

VERSION BE603735.2 GI:13187835

KEYWORDS EST.

SOURCE barley.

ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 955)
 Weng,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
 ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
 ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
 Wood,T.
 Development of a genetically and physically anchored EST resource
 for barley genomics
 Unpublished (2000)
 On Aug 21, 2000 this sequence version replaced gi:9861296.

TITLE

JOURNAL

COMMENT

Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: AATTAACCCCTCACTAAAGG
 High quality sequence stop: 931.

FEATURES

source

1. 955
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone_lib="HVSMEH0087003f"
 /HVCNA0009 (5 to 45 DAP)"
 /tissue_type="5-45 DAP Spike"
 /lab_host="SOLR"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 303 a 205 c 213 g 220 t 14 others

ORIGIN

Query Match 4.5%; Score 35.4; DB 137; Length 955;
 Best Local Similarity 52.0%; Pred. No. 5.7;
 Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 459 cgtgtttccaaatctcggactcaagccgactggttc 498
 Db 754 TTAGATGGCAAGCGTCCAGATGAAGTCCGCCGGTTC 793

RESULT 11
 A0746046 848 bp DNA GSS 16-JUL-1999
 LOCUS HS_2277_A2_F11_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2277 Col=22 Row=K, DNA sequence.

ACCESSION A0746046
 VERSION A0746046.1 GI:5523568
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 848)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2277 row: K column: 22
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 848.

FEATURES
 source
 1..848
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=2277 Col=22 Row=K"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli Dh10B"

BASE COUNT 228 a 136 c 117 g 144 t 223 others
 ORIGIN

Query Match 4.5%; Score 35.2; DB 233; Length 848;
 Best Local Similarity 55.8%; Pred. No. 6.4;
 Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 622 gttagaacagggcattcagtcggacaaatacgaaccccttagttatcttgggttcgac 681
 Db 334 GTTCTGCCAGGGCAATCAGGTAGGAGAAAGAAATCAACGGTATTCAATTAGGAAAGAG 393

Qy 682 catgtgcatgtactcgtatttgcaatgtcctgctgattattgaaatcagcaggac 741
 Db 394 GAAGTCAAAATGTCCCTGTTTGCAGATGACATGATTGTATATTAGAAACCCCATTTGTC 453

RESULT 12
 A0418525 487 bp DNA GSS 23-MAR-1999
 LOCUS RPCI-11-20318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-20318, DNA sequence.

ACCESSION A0418525
 VERSION A0418525.1 GI:4476249
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 555)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 487)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL Map Building
 COMMENT Unpublished (1997)
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbest@ig.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source
 1..487
 /organism="Homo sapiens"
 /db_xref="GB:7577767"
 /db_xref="taxon:9606"
 /clone="RPCI-11-20318"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT 165 a 102 c 96 g 124 t
 ORIGIN

Query Match 4.4%; Score 34.6; DB 228; Length 487;
 Best Local Similarity 54.3%; Pred. No. 8.4;
 Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 608 cctgttagaattgttaagaacagggcattcagtcggacaaatacgaaccccttagtt 667
 Db 127 CTCTGTGTAATTCGGTCCGTCAGGGCAATCAGGAGGAGAAATAAAGGGTGTTC 186

Qy 668 atcttggttcgacatgtcattctactcgtatttgccaatgtcctgctgattattga 727
 Db 187 AATTAGGAAAAGAGGAAGTCAATGTCCCTGTTTCAGATGACATGATTGTGTCTAC 246

Qy 728 aaatcagca 736
 Db 247 AAAACCCCA 255

RESULT 13
 A0417598 555 bp DNA GSS 23-MAR-1999
 LOCUS RPCI-11-203C8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203C8, DNA sequence.

ACCESSION A0417598
 VERSION A0417598.1 GI:4475441
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 555)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Thu Oct 25 13:08:00 2001

/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 126 a 260 c 245 g 157 t
ORIGIN

Query Match 4.4%; Score 34.4; DB 174; Length 788;
Best Local Similarity 52.0%; Pred. No. 11;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 55 cacttgaagacccggtcggcgactgcaaggattgaagactgttcttgggaatgacaggatt 114
DB 566 CCCAGGAGACAGAGGCTGGACAGCATGGAGCTGCAGGTGATCAGTCAATGACATTGTC 507
QY 115 caccctggagccgacccggtgttccgacgaacagaatagagactactgttcgaaggacgg 174
DB 506 CAGCTGCTGACGACCGGACCGCCGGCCACCAAGTGCACCGGCTCCAGCAGGAGATGAC 447
QY 175 gtgcttctcgagcacggagtcgccgactc 202
DB 446 GGCCGTCTCCAGCACGTTGCTCCCGCTC 419

Search completed: October 24, 2001, 13:13:45
Job time: 11810 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:20 ; Search time 393.52 Seconds
(without alignments)
930.236 Million cell updates/sec

Title: US-09-462-955-1_COPY_409_991
Perfect score: 583
Sequence: 1 actgctcttgagtcacaaag.....ttaaagtatgtgcatctaa 583
Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116.4	20.0	1096	20	AAV71833
2	116.4	20.0	1096	21	AAV71833
3	114.8	19.7	1091	20	AAV71834
4	114.8	19.7	1091	21	AAV71834
5	114.4	19.6	886	20	AAV71832
6	102.2	17.5	1106	20	AAV71832
7	100.6	17.3	1106	21	AAV71832
8	99.2	17.0	1017	17	AAV13165
9	91.2	15.6	1022	17	AAV13161
10	80.4	13.8	593	20	AAV71833
11	62.8	10.8	1110	19	AAV24089

12	59.6	10.2	982	18	AAV71833	Banana bunchy top
13	59.6	10.2	1110	19	AAV24086	Banana bunchy top
14	59.6	10.2	1111	19	AAV24077	Banana bunchy top
15	59.6	10.2	1111	19	AAV24084	Banana bunchy top
16	59.6	10.2	1111	19	AAV24087	Banana bunchy top
17	58.6	10.1	1103	19	AAV24091	Banana bunchy top
18	58.6	10.1	1105	19	AAV24093	Banana bunchy top
19	58	9.9	1109	19	AAV24088	Banana bunchy top
20	58	9.9	1111	19	AAV24090	Banana bunchy top
21	57	9.7	1104	19	AAV24092	Banana bunchy top
22	56.4	9.7	1111	19	AAV24085	Banana bunchy top
23	46.4	8.0	287	20	AAV71830	Subgenomic fragmen
24	46.4	8.0	287	21	AAV71831	Subgenomic fragmen
25	45.8	7.9	936	22	AAF58252	Oligonucleotide D1
26	45.8	7.9	936	22	AAF58254	Oligonucleotide D1
27	45.8	7.9	936	22	AAF58257	Oligonucleotide D1
28	45.8	7.9	936	22	AAF58259	Oligonucleotide D1
29	45.8	7.9	936	22	AAF58262	Oligonucleotide D2
30	45.8	7.9	938	22	AAF58255	Oligonucleotide D1
31	44.2	7.6	300	20	AAV71831	Subgenomic fragmen
32	44.2	7.6	300	21	AAV71831	Subgenomic fragmen
33	42.4	7.3	936	22	AAF58252	Oligonucleotide D1
34	42.4	7.3	936	22	AAF58254	Oligonucleotide D1
35	42.4	7.3	936	22	AAF58257	Oligonucleotide D1
36	42.4	7.3	936	22	AAF58259	Oligonucleotide D1
37	42.4	7.3	936	22	AAF58262	Oligonucleotide D2
38	42.4	7.3	938	22	AAF58255	Oligonucleotide D1
39	37	6.3	244	22	AAF58238	Oligonucleotide D1
40	34.2	5.9	244	22	AAF58238	Oligonucleotide D1
41	32.6	5.6	5053	19	AAV22816	Nucleotide sequenc
42	32.6	5.6	5055	19	AAV17236	DNA from a region
43	31.4	5.4	41599	21	AAZ35351	Cosmid including s
44	31	5.3	1212	20	AAV06636	Bacillus anthracis
45	31	5.3	1212	20	AAV06645	Bacillus anthracis

ALIGNMENTS

RESULT 1
AAV71833
ID AAV71833 standard; DNA; 1096 BP.
XX
AC AAV71833;
XX
DT 10-FEB-1999 (first entry)
XX
DE BTV DNA II clone (2-17) nucleotide sequence.
XX
KW Banana bunchy top virus; BTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
XX
OS Banana bunchy top virus.
XX
FH Key Location/Qualifiers
FT TATA_signal 1..7
FT /*tag= a
FT stem_loop 8..38
FT /*tag= b
FT CDS 70..928
FT /*tag= c
FT /*product= "ORF-V2 product"
FT /*transl_except= (pos:215..217, aa:Gly)
FT polyA_signal 533..538
FT /*tag= d
FT polyA_signal 795..804
FT /*tag= e
FT polyA_signal 907..912
FT /*tag= f
FT polyA_signal 1030..1035
FT /*tag= ge
XX
PN US5846705-A.

[illegible]

RESULT	3
AAV71834	

ID AAV71834 standard; DNÄ; 1091 BP.

XX
AC AAV71834:

DT 10-FEB-1999 (first entry)

DE BBTV DNA II clone (2) nucleotide sequence.

XX Banana bunchy top virus; BRTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
XX
OS Banana bunchy top virus.

XX
PN
US5846705-A.

08-DEC-1998.

XX
PF 06-APR-1995; 95US-0418071.

XX
PR 06-APR-1995; 95US-0418071.

XX
PA (BIOT-) DEV CENT BIOTECHNOLOGY.

XX
PI Soong T, Wu R, You L;

XX
DR WPI; 1999-059037/05.

Nucleic acids having banana bunchy top virus component sequences used to design primers for use in polymerase chain reaction detection of the virus

Disclosure; Fig 12A-C; 27pp; English.

This represents the nucleotide sequence of a banana bunchy top virus (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention provides nucleic acid sequences associated with BBTV that can be used in a PCR technique for detecting BBTV. The nucleic acid sequences (AAV71030 to AAV71833) are used as the basis for the construction of PCR primers, to detect BBTV infection. The PCR technique is used for detecting BBTV in plant tissues (preferably banana, especially *Musa* species). The virus, one of the most important banana species viruses, causes phloem damage and is transmitted by aphids. PCR detection gives accurate, reliable and specific determination of absence or presence of the virus.

Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other;

Query Match

Query Match 19.7%; Score 114.8; DB 20; Length 1091;
Best local similarity: 50.0%

Best Local Similarity 52.6%; Pred. No. 1.3e-28;

Matches	304;	Conservative	0;	Mismatches	262;	Indels	12;	Gaps	2;	
QY	6	tcttgagtc	aaagccacg	attggcccaacg	atttctgaggaac	ctgatgaactcg	65			
Db	368	tcttggtcga	agcgcaagctt	ctcgatagatcgagagag	acctcgaggaa	ttagaa	427			
QY	66	cttggaagac	caggcgat	accgaagatcggtgtacacg	agcttcggtgg	aatggac	125			
Db	428	gatggacgt	atccatccag	atctcgagatgcttggcag	tggaatca	atgaagatgccag	487			
QY	126	aagatggcgcg	ctgaaatcgcgttcc	atttccatatacac	aatggcagcttga	agtgt	185			
Db	488	aattaatccg	atggtgttcga	aacta-----	aaagaatgg	caaaaataa	taata	538		
QY	186	gtctcgatc	ggagagccg	cgagcgtatcgacac	aatctctggtatcg	gcgacagacgg	245			
Db	539	tcacacatc	gaaggtgttct	gatgatcggaagtat	catcttggtatac	ggtctcctgcgg	598			
QY	246	aggacacggg	aagtcggtgttgc	aaatactcggact	caagccgactggtt	ctctacac	305			
Db	599	aggcgaag	aagtcac	acttcgaagatactat	cattaaa	acctggatggg	gatata	658		
QY	306	atgtgtgtg	accacgaag	gcgtatgtaccg	tacatcgagagacc	caaaacga	aat	365		
Db	659	caacggtgga	agcgtcggtatg	atgcacatcaac	gatggatcgtata	atcattg	718			
QY	366	aatctcgat	gtacc	agggtgataatt	tagagattt	aaattatgcctgt	tagaatgt	425		
Db	719	gattatgtat	ccccgaag	ctcattccag	attctg	aatattg	cggtatag	aca	aat	778
QY	426	taagaacagg	ctatc	agttcggac	aaatacga	ac-----	collag	tattcttgggttcga	482	
Db	779	taagaatag	agtttga	ataaataca	aaatacga	acctgtgtatt	agaaagatg	gaca	838	
QY	483	ccatgtcat	gtactgtattt	gccaatgtcctgcctgatt	tcttgaa	aatcagc	agga	542		
Db	839	aaatgtccat	gtaatgttat	ggcaaatgtgtgcctgatt	gtataa	aatcttcaga	aaga	898		
QY	543	cagaataa	aaactgtgg	aatattt	aaagtatgt	gcattc	580			
Db	899	tagataa	aaataa	ttaattgtt	ggaaagga	aaacttc	936			

RESULT

RESULT
AAA38947

AAA38947
ID AAA38947 standard; DNA; 1091 BP.

XX
AC

XX
DT 23-AUG-2000 (first entry)

XX
DE
XX

XX
KW
Banana bunchy top virus; BBTV; detection; ds.

XX Banana bunchy top virus.

XX
PN TW360710-A.

XX
PD 11-JUN-1999.

XX
PF 30-JUN-1994; 94TW-0106105.

XX
PR 30-JUN-1994; 94TW-0106105.

XX
PA (BIOT-) DEV CENT BIOTECHNOLOGY.

XX
PI Wu R, You L, Song T;

XX
DR WPI; 2000-316145/27.

XX
PT- Two circular single-stranded DNAs associated with banana bunchy top
PT virus and detection of the virus -

XX Claim 2; Page 1; 7pp; Chinese.

XX The present invention describes two circular single-stranded DNAs

CC associated with banana bunchy top virus (BBTV) and the encoded proteins.

CC Also described is a method of detecting the BBTV virus using the

CC polymerase chain reaction (PCR). The present sequence represents a

CC BBTV related nucleotide sequence from the present invention.

XX Sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;

XX

Query Match 19.7%; Score 114.8; DB 21; Length 1091;

Best Local Similarity 52.6%; Pred. No. 1.3e-28;

Matches 304; Conservative 0; Mismatches 262; Indels 12; Gaps 2;

QY 6 tctgagatcaaaagccacgattggcccaacgattgtcgaggaaacctgatgaactcgg 65

DB 368 tctggttcgaagagcgcaagctcttcgataagattcagagagagccctgaagaattgaa 427

QY 66 cctggaagaccagcggtatccggaagatcgctgtgtacacggagcttcggtggaatggac 125

DB 428 gatgacgatccatccaagtatcgagatcgctgtgagtggaatcaattaaagatgccag 487

QY 126 asgatggccgctgaaatccgcttcctccatccatatacaaatggcagctgaagtgc 185

DB 488 aaataattccgaatgggttcacgaacta-----aaagaatggcacaataaattaat 538

QY 186 gtcctgcatcgagagcgagcgagcgtatcgacacatctctggtatgcgagcagacgg 245

DB 539 tcaacacatcgaggtgtctgctgatcgagatcatctcgggtatcagcgtctgcgcgg 598

QY 246 aggagacgggaagtcgctgtgttgcgaatatctcggacacacgacgactggtctcac 305

DB 599 aggcgaagaaagtcacaccttcgcaagatatctattcaaaacctggatggggatata 658

QY 306 atgtgtggaaccagaaagcgatattgtacacgacgacgacgacgacgacgacgacgac 365

DB 659 caacggtggaagaaagcgatattgtacacgacgacgacgacgacgacgacgacgacgac 718

QY 366 aatcctcgatgacccaggtgtaatttagagattttaaattatgacctgttagaattgt 425

DB 719 gattattgatcccgacagagcattcagattatctggaattatgacctgttagaacaat 778

QY 426 taagaacagggcattcagttcggacaaatacgaacc---ccttagttactgtggttcga 482

DB 779 taagaatagagttttaaataaatacaaaatacgaaccatgtgtgattagaagaatggaca 838

QY 483 ccatgtgcatgactgctgatttgcgaatgctcgtgctgattatttgaataatcagcaggg 542

DB 839 aatgtccatgtaattgtttatggcaaatgtgtgctgattatttgaataatcagcaggg 898

QY 543 cagaataaaacttggaattattaaagtatgctatc 580

DB 899 tagaataaaataattattgttgagaaggaacttc 936

RESULT 5

AAV71832

ID AAV71832 standard; DNA; 886 BP.

XX

AC AAX34687;

XX

DT 02-JUL-1999 (first entry)

XX DNA sequence of BBTV S21 and S2L promoter fragments.

DE

XX Promoter; Banana bunchy top virus; BBTV; gene transcription; ubil;

XX polyubiquitin 1; banana; ss.

XX

OS Banana bunchy top virus.

XX

PN WO9915646-A1.

XX

PD 01-APR-1999.

XX

PF 21-SEP-1998; 98WO-AU00786.

XX

PR 30-JUN-1998; 98AU-0004423.

PR 19-SEP-1997; 97AU-0009339.

XX (YUQU-) UNIV QUEENSLAND TECHNOLOGY.

XX

XX Becker DK, Dale JL, Dugdale B, Harding RM, Hermann SR;

XX

XX WPI; 1999-254706/21.

XX

XX Banana bunchy top virus promoter constructs

XX

XX Examples: Fig 23; 84pp; English.

XX The invention relates to DNA promoter sequences derived from components

CC of Banana bunchy top virus (BBTV). The DNA molecule includes a promoter

CC sequence derived from an untranslated portion of any one BBTV components

CC and is adaptable for promoting transcription of a cloned gene in a plant

CC cell. The promoter, included in a DNA chimeric vector, is useful for the

CC expression of a gene in a plant cell. The inclusion of the polyubiquitin

CC 1 (ubil) intron sequence into BBTV promoter constructs enhances promoter

CC activity. The present sequence represents the DNA sequence of BBTV S21

CC promoter and S2L promoter fragments.

XX

XX Sequence 886 BP; 294 A; 171 C; 199 G; 222 T; 0 other;

XX

Query Match 19.6%; Score 114.4; DB 20; Length 886;

Best Local Similarity 56.9%; Pred. No. 1.6e-28;

Matches 230; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 167 attgcagcttgaaagtgtctgctgagcagagagcagcagcagcagcagcagcagcagc 226

DB 246 aatggcaaaataataataataataataataataataataataataataataataataata 305

QY 227 ggatgctggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 286

DB 306 ggtatatacgttcccaacgagcagcagcagcagcagcagcagcagcagcagcagcagc 365

QY 287 agccagcagctgttcaacatgtgtgcaacgagcagcagcagcagcagcagcagcagcag 346

DB 366 aacctgtgtgggataatacaacgctggagagcagcagcagcagcagcagcagcagcagc 425

QY 347 aggaccccaaaacgaaatttaattcctcgtatgacccagctgtaatttagagatttaatt 406

DB 426 tggatcctgataatcatgttgatttattgataccccaagcagcagcagcagcagcagcag 485

QY 407 atgcccctgttagaattgttgaagaacagcagcagcagcagcagcagcagcagcagcag 463

DB 486 atggcgttatagaacaaataagaatagatttttaataatacaaaatacaaaacacatgtg 545

QY 464 ttatttattctgtgttcgacctgtgctgctgctgctgctgctgctgctgctgctgctg 523

DB 546 tgattagaagaagatggacaaatgcccattgttattgttattgttattgttattgttatt 605

QY 524 atttgaataatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 567

DB 606 attgtaaaatttcagaagataagaataaaaaataaattattgttga 649

RESULT 6

AAV71832

ID AAV71832 standard; DNA; 1106 BP.

XX

AC AAV71832;

XX

DT 10-FEB-1999 (first entry)

XX BBTV DNA I clone (7-4-2) nucleotide sequence.

DE

Query Match 17.0%; Score 99.2; DB 17; Length 1017;
Best Local Similarity 54.1%; Pred. No. 2.4e-23;
Matches 225; Conservative 0; Mismatches 188; Indels 3; Gaps 1;

```

163 cacaatggcagcttgaagtgcgtcgatcgaggagccgacgagcatgcacaaac 222
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 498 cgccctggcaagaagatgatcttgtagaagaaaccagattatagaacgata 557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
223 ctctggatatgcgcacagacgagcggagacggaaactccgttttgcacaaatctcggg 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 558 atctgggtgtatggacctgctggaatgaaggccaatctcacattfccaagacatctgtca 617
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 ctcaagcccactggtttctcacacatggtgtggaaccagaaaaggacgtattgtaccagttac 342
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 618 tgaagaagtgttggtttatcttcctcgtggagaaagacacaagatatgatgcatttgtg 677
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
343 atcgaggaccaccaaacgaatttaactcctcgatgtaccagggtgtaatttagagtattta 402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 678 actgctgagcctaagaataatgttggttatctgacataccccagagttagttcagagtatgtg 737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 403 aattatgccctgttagaatgtttaagaacagggcattcagttcggacaaaacacgaacc- 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 738 aattatgtgttaatgaacaggtttaagaatagggttaattgtgaataactaatgatgagcca 797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 462 --cctagtattccttgggttcgacatgtgcattctactcgtatttgcacaaatgcctgct 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 798 tgttaatgcgggatgaataatcctctgttcattgtaattgtttgcaaatgtactccca 857
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 520 gattattgaaataacgacgagcagacagataaaactctggaataatttaaagtatgtg 575
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 858 gatttggaaaaataatgaatgaagatagataaaaaataattcgttgttgaaactctg 913
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 9
AAU13161
ID AAU13161 standard; DNA; 1022 BP.
AC AAU13161;
XX XX
DT 23-MAY-1996 (first entry)
XX XX
DE SCSV segment 2.
XX XX
KW SCSV; promoter; transcription; transgenic plant; legume;
XX gene expression; crop improvement; SS.
OS Subterranean clover stunt virus isolate F.
XX XX
FH Location/Qualifiers
FT CDS 82..924
FT /*tag= a
XX XX
PN WO9606932-Al.
XX XX
PD 07-MAR-1996.
XX XX
PF 30-AUG-1995; 95WO-AU00552.
XX XX
PR 07-NOV-1994; 94AU-0009281.
PR 30-AUG-1994; 94AU-0007770.
XX XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA (AUSU) UNIV AUSTRALIAN NAT.
XX XX
PI Boevink PC, Chu PWG, Keese PK, Khan RI, Larkin PJ;
PI Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
XX XX
DR WPI; 1996-160363/16.
XX XX
PT Circovirus transcription regulatory sequences and related constructs
PT - useful in plants, esp. leguminous plants, for the modulation of

PI Becker DK, Beetham PR, Dale JL, Dugdale B, Hafner GJ;
PI Harding RM;
XX WPI; 1997-034368/03.
XX
XX DNA from intergenic region of banana bunchy top virus DNA component
PT - useful for promoting, enhancing, regulating or modifying
PT transcription of a non-BBTV gene in transgenic plants
XX
XX Claim 4; Fig 12; 80pp; English.

XX Intergenic regions (AA749399-409) are derived from genomic DNA
CC components 1-6 (see also AA749386-90) of banana bunchy top virus
CC (BBTV). A large intergenic region (AA749405) comprises the insert
CC in plasmid pBT1.1NT, a plasmid obtd. by PCR amplification
CC (see also AA749427-28) of a full-length BBTV1 clone and cloning of
CC TaqI-digested PCR product into pBT101.3. BBTV intergenic regions
CC are useful for promoting, enhancing, regulating or modifying
CC transcription of non-BBTV genes in monocotyledon or dicotyledon
CC transgenic plants. The non-BBTV gene is e.g. an insecticide
CC resistance gene, herbicide resistance gene or growth promoting
CC gene.

XX SQ Sequence 982 BP; 316 A; 174 C; 249 G; 243 T; 0 other;

Query Match 10.2%; Score 59.6; DB 18; Length 982;
Best Local Similarity 51.3%; Pred. No. 6e-10;
Matches 193; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

QY 180 agtgcgtctgcgtatcgagagccagcgagcgatcgacacatctctggatcgagcag 239
DB 53 agagtgaggaataatcgccgagccatcgatcgagagataatttggctctatggccc 112
QY 240 agacgagagagcgagagcgagtcgtgttgcacaaatctctggactcaagcccgactggtt 299
DB 113 aaatggagagagagagagagacacgtagtgcacaaatctctgaagcgagaaatcggtt 172
QY 300 ctacacatgtgtggaacacagagagagcgatgtgtaccagtagacgagagacacaaacg 359
DB 173 ttattctccagat 229
QY 360 aaatttaattctgcgtatcgagagcgagtcgtgttgcacaaatctctgaagcgagaaatcggtt 419
DB 230 ---tgttatatttgatattccagagatgcagagagagagagagagagagagagagagagag 286
QY 420 atgtgttaagagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 479
DB 287 ggaatttaag 343
QY 480 cgacatgtgcgtatcgagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 539
DB 344 agaatatgcgaagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgag 403
QY 540 ggaacag 555
DB 404 agatcgaataaagttg 419

RESULT 13
AAV24086
ID AAV24086 standard; DNA; 1110 BP.
XX
XX AC AAV24086;
XX
XX DT 11-AUG-1998 (first entry)
XX DE Banana bunchy top virus component 1 DNA sequence.
XX KW BBTV; probe; diagnostic primer; component 1; ss.
XX OS Banana bunchy top virus.
XX

PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
PI Burns TM, Dale JL, Harding RM, Karan M;
XX WPI; 1998-321636/28.
XX
XX Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers and for producing virus-resistant plants
XX
XX Disclosure; Column 23-26; 59pp; English.

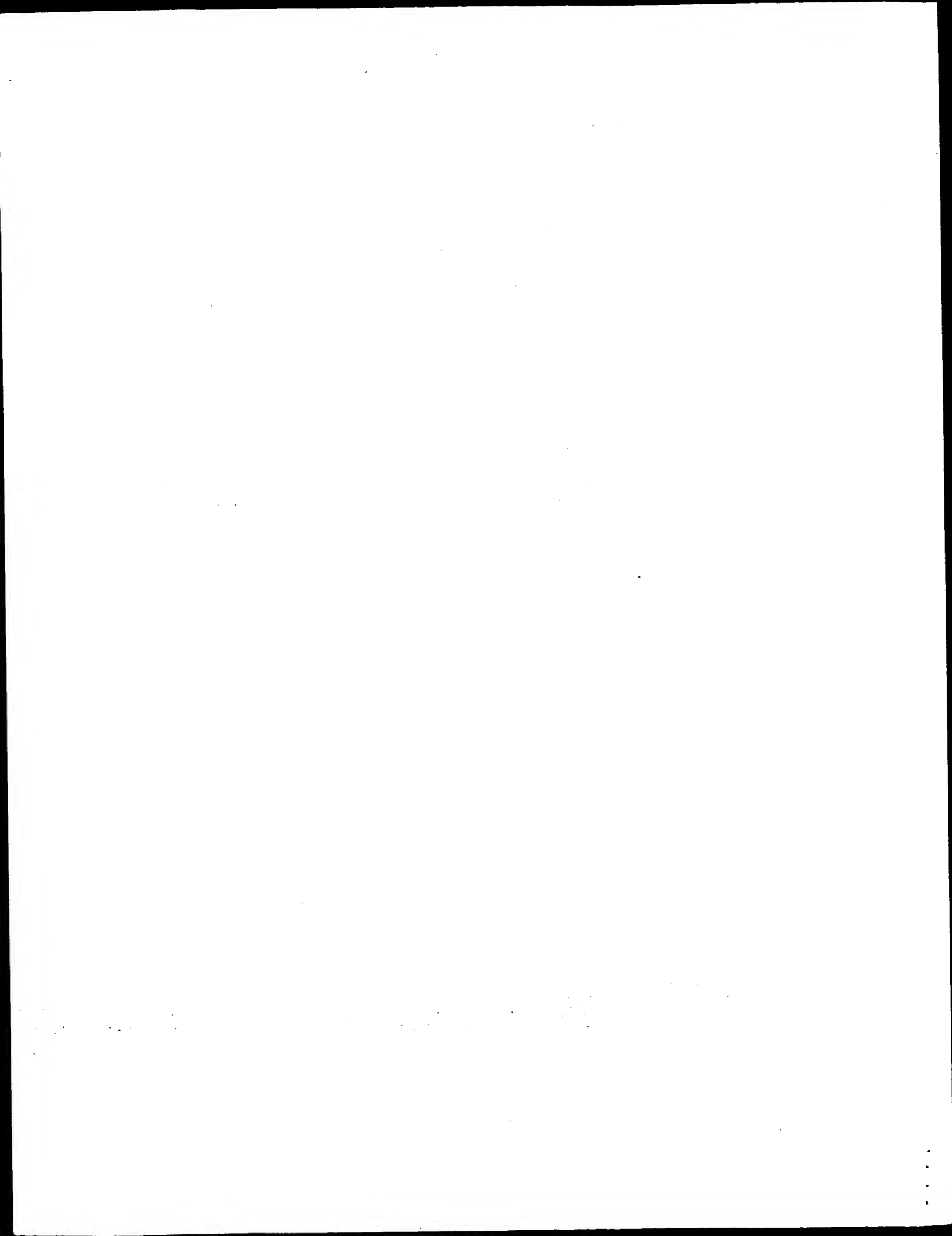
XX This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plants or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
CC enhancers or termination signals.

XX SQ sequence 1110 BP; 358 A; 197 C; 275 G; 280 T; 0 other;

Query Match 10.8%; Score 62.8; DB 19; Length 1110;
Best Local Similarity 53.2%; Pred. No. 5.3e-11;
Matches 183; Conservative 0; Mismatches 152; Indels 9; Gaps 2;

QY 212 atcgacacatctctgcgtatcgagagcgagcgagcgagcgagcgagcgagcgagcgagcgag 271
DB 615 atcgagaattatttgggttacgcccacaaatggagagagagagagagagagagagagagagag 674
QY 272 aatatctgcgactaacgcccgcgtgttctacacatgtgtggaacacagagagcgat 331
DB 675 aacatctaataagacgat 734
QY 332 tgtaccagtagacatcgagagacacacaaacgaatttaactctgcgtatcgagcgagcgag 391
DB 735 gttagctgataattatgagat-----tgttatatttgatctccctagatgcaag 788
QY 392 tagagctatttaattatgccctgttagaattgtttaagacagcgagcgagcgagcgagcgag 451
DB 789 aggatatttaattatggtttatgagagagagagagagagagagagagagagagagagagag 848
QY 452 aatacgaaccccttagtattctgtggttcgacacatgtgcacatgtgcgtatcgagcgagcgag 511
DB 849 aatatgaacccgtttgagagattg---tagaataatgcgaagtcgagcgagcgagcgagcgag 905
QY 512 tctgctgattatttgaaataacagcagcgagcgagcgagcgagcgagcgagcgagcgagcgag 555
DB 906 tcttccgaaggaagaatcttttctgagagatcgagcgagcgagcgagcgagcgagcgagcgag 949

RESULT 12
AA749405
ID AAT49405 standard; DNA; 982 BP.
XX
XX AC AAT49405;
XX
XX DT 23-AUG-1997 (first entry)
XX DE Banana bunchy top virus DNA intergenic region 1 insert in pBT1.1NT.
XX KW BBTV; intergenic region; promoter; transgenic plant; ds.
XX OS Banana bunchy top virus.
XX PN WO9638554-A1.
XX PD 05-DEC-1996.
XX PF 31-MAY-1996; 96WO-AU00335.
XX PR 31-MAY-1995; 95AU-0003285.
XX PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
XX



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:45 ; Search time 5479.82 seconds
(without alignments)
1005.692 Million cell updates/sec

Title: US-09-462-955-1_COPY_409_991

Perfect score: 583

Sequence: 1 actgcctcgtgagtcacaaag.....ttaaagtatgtgcacataa 583

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_estl1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est33:*
- 26: gb_est34:*
- 27: gb_est35:*
- 28: gb_est36:*
- 29: gb_est37:*
- 30: gb_est38:*
- 31: gb_est39:*
- 32: gb_est40:*
- 33: em_estba:*
- 34: em_estfun:*
- 35: em_esthum1:*
- 36: em_esthum2:*
- 37: em_esthum3:*
- 38: em_esthum4:*
- 39: em_esthum5:*
- 40: em_esthum6:*
- 41: em_esthum7:*
- 42: em_esthum8:*
- 43: em_esthum9:*
- 44: em_esthum10:*
- 45: em_esthum11:*
- 46: em_esthum12:*
- 47: em_esthum13:*
- 48: em_esthum14:*
- 49: em_esthum15:*
- 50: em_esthum16:*
- 51: em_esthum17:*
- 52: em_esthum18:*
- 53: em_esthum19:*
- 54: em_esthum20:*
- 55: em_esthum21:*
- 56: em_esthum22:*
- 57: em_esthum23:*
- 58: em_esthum24:*
- 59: em_esthum25:*
- 60: em_esthum26:*
- 61: em_esthum27:*
- 62: em_esthum28:*
- 63: em_estin1:*
- 64: em_estin2:*
- 65: em_estin3:*
- 66: em_estin4:*
- 67: em_estin5:*
- 68: em_estom1:*
- 69: em_estom2:*
- 70: em_estov1:*
- 71: em_estov2:*
- 72: em_estpl1:*
- 73: em_estpl2:*
- 74: em_estpl3:*
- 75: em_estpl4:*
- 76: em_estpl5:*
- 77: em_estpl6:*
- 78: em_estpl7:*
- 79: em_estpl8:*
- 80: em_estpl9:*
- 81: em_estpl10:*
- 82: em_estro1:*
- 83: em_estro2:*
- 84: em_estro3:*
- 85: em_estro4:*
- 86: em_estro5:*
- 87: em_estro6:*
- 88: em_estro7:*
- 89: em_estro8:*
- 90: em_estro9:*
- 91: em_estro10:*
- 92: em_estro11:*
- 93: em_estro12:*
- 94: em_estro13:*
- 95: em_estro14:*
- 96: em_estro15:*
- 97: em_estro16:*
- 98: em_estro17:*
- 99: em_estro18:*
- 100: em_estro19:*
- 101: em_estro20:*
- 102: gb_est25:*
- 103: gb_est26:*
- 104: gb_est27:*
- 105: gb_est28:*
- 106: gb_est29:*
- 107: gb_est30:*
- 108: gb_est31:*
- 109: gb_est32:*
- 110: gb_est41:*
- 111: gb_est42:*
- 112: gb_est43:*
- 113: gb_est44:*
- 114: gb_est45:*
- 115: gb_est46:*
- 116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

/clone_lib="Barstead mouse myotubes MPLRB5"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with Not I - oligo(dT) primer [5',
TGTATCAATCTGAAGTGGACGGCGCCCTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins. "
117 a 126 c 154 g 118 t 1 others
BASE COUNT

```

[illegible]

RESULT	4
AQ615572	DNA
LOCUS	498 bp
DEFINITION	HS_5144_B1_B02_SP6E_RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=720 Col=3 Row=D, DNA sequence.
ACCESSION	AQ615572
VERSION	AQ615572.1
KEYWORDS	GI:5076848
SOURCE	GSS.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 498)
AUTHORS	Mahdairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and

JOURNAL	MEDLINE	COMMENT	FEATURES
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	99300589	<p>Contract: Mahiras GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu</p> <p>Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu</p> <p>Plate: 720 row: D column: 3 Seq primer: SP6 Class: BAC ends High quality sequence stop: 498. Location/Qualifiers</p>	

source

1. .498
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-720 Col-3 Row-D"
 /clone_lib="RPC1-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACE3.6 vector at EcoRI sites"
 BASE COUNT 171 a 107 c 87 g 113 t 20 others
 ORIGIN

Query Match 6.3%; Score 36.8; DB 231; Length 498;
 Best Local Similarity 56.7%; Pred. No. 1.5;
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
 QY 424 gtaagaacaggcattcagtcgacaaatacgaaccccttagttatcttggttcgac 483
 DB 271 GTTCTGGCCAGGCAATCAGSCAAGCAAGAAATAAGGGTATTAGTTAGGAAGAG 330
 QY 484 catgtcgtactcgtatttgcaatgcctcgtcgtgatttgaataacagcaggac 543
 DB 331 GAAGTCAAAATGTCCCTGTTGCGAGATGACATGATTGTTTATTAGAAATCCCATCGAC 390

RESULT 5

AW818171 518 bp mRNA EST 17-MAY-2000
 LOCUS CM1-ST0277-161299-070-h05 ST0277 Homo sapiens cDNA, mRNA sequence.
 DEFINITION AW818171
 ACCESSION AW818171
 VERSION AW818171.1 GI:7911165
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1. (bases 1 to 518)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM1-ST0277-161>)
 299-070-h05&t3=1999-12-16&t4=1

Seq primer: puc 18 forward

High quality sequence start: 76

High quality sequence stop: 183.

FEATURES

Source

1. .518
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="ST0277"
 /dev_stage="Adult"
 /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products

BASE COUNT

ORIGIN

derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 193 a 119 c 89 g 117 t

Query Match

Best Local Similarity 61.5%; Score 36.8; DB 121; Length 518;
 Matches 59; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 347 aggaacacaaacgaatttaactcctcgtatccacagtgtaatttagatttaatt 406

DB 168 AGGATTAAAAACCAATTTGATCTCATGACACAGCAAAAATAATAGTACACAGATT 227

QY 407 atgcccgttagaattgtttaagaacagggcaltca 442

DB 228 GTGCTCAGAAATAACGTCTTAAACACCTGCCCAATCA 263

RESULT 6

AW889903/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

UNIVERSITY

ADDRESS

TELEPHONE

FAX

EMAIL

SEQUENCE

CLASS

HIGH QUALITY

SEQUENCE STOP

LOCATION/QUALIFIERS

/ORGANISM

/DB_XREF

/CLONE_LIB

/SEX

/NOTE

E-COLI DH10B

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

Thu Oct 25 13:08:12 2001

/clone="Plate-2172 Col=15 Row=N"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 172 a 91 c 72 g 107 t 1 others
ORIGIN

Query Match 5.8%; Score 34; DB 223; Length 443;
Best Local Similarity 55.7%; Pred. No. 11;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Qy 424 gttaaagacagggcattcagttcggacaaatcacgaaccccttagttatcttgggttcgac 483
Db 318 GTTTTGGCCAGGCGAGTCAAGGAGAGAAATAAAAGCGTATTCAATAAGGAGAGAA 377
Qy 484 catgtgcatgtactcgtatttgccaatgtcctgcctgatttattgaaatcagca 538
Db 378 GAAGTCAAATGCCCATGTTTGCAGATGACATGATGTTTATNTAGAAAACCCCA 432

Search completed: October 24, 2001, 13:13:46
Job time: 11811 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:23 ; Search time 180.6 seconds
(without alignments)
611.121 Million cell updates/sec

Title: US-09-462-955-1_COPY_409_991

Perfect score: 583
Sequence: 1 actcgtcctggagtcacaaag.....ttaaagtatgtcatctaa 583

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6C_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/6D_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116.4	20.0	1096	2	US-08-418-071-4
2	114.8	19.7	1091	2	US-08-418-071-5
3	102.2	17.5	1106	2	US-08-418-071-3
4	99.2	17.0	1017	4	US-08-793-634B-6
5	91.2	15.6	1022	4	US-08-793-634B-2
6	62.8	10.8	1110	1	US-08-202-186-14
7	59.6	10.2	982	3	US-08-973-068-28
8	59.6	10.2	1110	1	US-08-202-186-11
9	59.6	10.2	1111	1	US-08-202-186-9
10	59.6	10.2	1111	1	US-08-202-186-12
11	59.6	10.2	1111	1	US-08-202-186-24
12	58.6	10.1	1103	1	US-08-202-186-16
13	58.6	10.1	1105	1	US-08-202-186-18
14	58	9.9	1109	1	US-08-202-186-13
15	58	9.9	1111	1	US-08-202-186-15
16	57	9.8	1104	1	US-08-202-186-17
17	56.4	9.7	1111	1	US-08-202-186-10
18	46.4	8.0	287	2	US-08-418-071-1
19	44.2	7.6	300	2	US-08-418-071-2
20	32.2	5.5	1218	3	US-08-232-463-14
21	31	5.3	1212	3	US-09-046-578-5
22	30.4	5.2	1212	3	US-09-046-578-1
23	29.4	5.0	8920	4	US-08-446-855A-1
24	29.4	5.0	8920	4	US-09-150-741-1
25	29.2	5.0	1209	6	5352575-4
26	28.8	4.9	10607	1	US-08-078-090-3
27	28.6	4.9	1593	2	US-08-524-828-2
					Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705

; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh

; APPLICANT: You, Li-Ru

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS

; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: United States

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 5.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/418,071

; FILING DATE: 06-APR-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Haley Jr., James F.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: DCB-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 596-9000

; TELEFAX: (212) 596-9090

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1096 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: circular

; MOLECULE TYPE: genomic DNA

; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Banana Bunchy Top Virus (BBTV)

US-08-418-071-4

Sequence 2, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 9, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 4, Appli

```

Query Match      20.0%; Score 116.4; DB 2; Length 1096;
Best Local Similarity 52.8%; Pred. No. 9.9e-30;
Matches 305; Conservative 0; Mismatches 261; Indels 12; Gaps 2;

QY 6 tcttgagtcacaaagccacagatggtcccaacgattgtgaggaacactgatgaactccg 65
Db 373 TCCTGGTTCCGAAGAAGCCCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 432
QY 66 cctgaagaccagcgcgataccgaagatggtgtacacgagcttcggtggaatggac 125
Db 433 GATGACGATCCATCCAGTATCGCAGATGCTGCGCATGGGAATCAATTAAGATGCCAG 492
QY 126 aagatggcgctgaaatccgttcccatccattccatcacaaattggcagctgaagtgt 185
Db 493 AATTAATTCGAATGGTTCCAGAACTA-----AAAGAATGGCAAAATAAATAAT 543
QY 186 gtctgcgacgcagagcagcgagcgcacacatctctggtatgcgagcagagcgg 245
Db 544 TCAACACATCGAAGGTGTCCTGATGATCGAAGTATCATCTCGGTATACGGTCCCAACGG 603
QY 246 aggagacgggaagtcggtgtttgccaataatctcgggaactcaagcccgactggtctacac 305
Db 604 AGCGGAAGAAAGTCAACCTTCGCAAGATATCTATCATTAACCCCGGATGGGGATATAT 663
QY 306 atgtggtgaaccagaaagacgattgtaccagttacatcgcagacaccccaaacgaattt 365
Db 664 CAACGCTGGAAGACGTCGGATGATGACATCATACATGATGATGATGATGATGATGATGAT 723
QY 366 aatctcgtgtaccaggtgtaattagatttttaataattatgcctgttagaattgt 425
Db 724 GATTATTGATATCCCGAGAAGTCAATTCAGATATCTGAATTTATGCGGTATAGAACAAAT 783
QY 426 taagaacagggcattcagtcgcgacaaatacgaacac-----ccttagttatcttggttcca 482
Db 784 TAAGAATAGAGTTTAAATAATACAAAATACGAACCATGTCATGTAGAAAAAGATGGACA 843
QY 483 ccatgtcgtactgctgttggcgaatgctgcctgctgattttgaaatcagcaggga 542
Db 844 AATGTCATGTAATTTATGTCGAATGTTGTCCTGATTATGTTGTTGTTGTTGTTGTTGTTGTT 903
QY 543 cagaataaaactggaataatttaagttatgtgtatc 580
Db 904 TAGAATAAAATAAATAATTTGTTGAGAAAAGGAACCTTC 941

RESULT 2
US-08-418-071-5
; Sequence 5, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTING BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: DCB-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: subgenomic DNA
DESCRIPTION: /desc "BBTV DNA II (clone 2)"
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-5

```

```

Query Match      19.7%; Score 114.8; DB 2; Length 1091;
Best Local Similarity 52.6%; Pred. No. 3.4e-29;
Matches 304; Conservative 0; Mismatches 262; Indels 12; Gaps 2;

QY 6 tcttgagtcacaaagccacgattggcccaacgattgtgaggaacactgatgaactccg 65
Db 368 TCCTGGTTCCGAAGAAGCCCAAGCTTCTCGATAGATTTCAGAGAGAGCCCTGAGGAATTGAA 427
QY 66 cctgaagaccagcgcgataccgaagatgctgtgtacacgagcgttcggtggaatggac 125
Db 428 GATGACGATCCATCCAGTATCGCAGATGCTTGGCATGGAATCAATTAAGATGCCAG 487
QY 126 aagatggcgctgaaatccgttcccatccattccatcacaaattggcagctgaagtgt 185
Db 488 AATTAATTCGAATGGTTCCAGAACTA-----AAAGAATGGCAAAATAAATAAT 538
QY 186 gtctgcgacgcagagcagcgagcgcacacatctcctgtggtatgcgagcagagcgg 245
Db 539 TCAACACATCGAAGGTGTTCTGATGATCGAAGTATCATCTGGGTATACGGTCCCTGCCGG 598
QY 246 aggagacgggaagtcggtgtttgccaataatctcgggaactcaagcccgactggtctacac 305
Db 599 AGCGGAAGAAAGTCAACCTTCGCAAGATATCTATCATTAACCCCGGATGGGGATATAT 658
QY 306 atgtggtgaaccagaaagacgattgtaccagttacatcgcagacaccccaaacgaattt 365
Db 659 CAACGCTGGAAGAGCTGCGGATATGATGCACATCATACGATGATGATGATGATGATGATGAT 718
QY 366 aatctcgtgtaccaggtgttaatttagatttttaataattatccctgttagaattgt 425
Db 719 GATTATTGATATCCCGAGAAGTCAATTCAGATATCTGAATTTATGCGGTATAGAACAAAT 778
QY 426 taagaacagggcattcagtcgcgacaaatacgaacac-----ccttagttatcttggttcca 482
Db 779 TAAGAATAGAGTTTAAATAATACAAAATACGAACCATGTCGATTAGAAAAAGATGGACA 838
QY 483 ccatgtcgtactgctgttggcgaatgctgcctgctgattttgaaatcagcaggga 542
Db 839 AATGTCATGTAATTTATGTCGAATGTTGTCCTGATTATGTTGTTGTTGTTGTTGTTGTTGTT 898
QY 543 cagaataaaactggaataatttaagttatgtgtatc 580
Db 899 TAGAATAAAATAAATAATTTGTTGAGAAAAGGAACCTTC 936

```

```

RESULT 3
US-08-418-071-3
; Sequence 3, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh

```


Query Match	17.5%	Score 102.2	DB 2	Length 1106
Best Local Similarity	53.3%	pred. No. 6.4e-25		
Matches 215	Conservative 0	Mismatches 188	Indels 0	Gaps 0
QY 169	tggcagcttgaa	gtctgtctgcgacatcgagagccagcgagcagatcgacacatactctcgg	228	
Db				
QY 524	TGGCAGATTCA	ATTGACGGAGCCAATTGACGAGGACCCGCGATCGACGAGCATCATCTGG	583	
Db				
QY 229	atatcgagacagag	cagagagacaggaadtcctgtgtttgccaaatatactcggaactcaag	288	
Db				
QY 584	GTCTATGGTC	CTTATGGTAAATGAGGTAAATCAACATATGCGAAGTCACTAATCAAGAAG	643	
Db				
QY 289	ccgcactggttc	atacatatgtgtggaacacagaaagacgtattgtaccagtaactcag	348	
Db				
QY 644	GATTGGTCTT	ACACGAGGGTGGAGAGGAGAAATATCTTATTCTCTATGTGGACGAA	703	
Db				
QY 349	gacccaaacgaa	attaatcctctgatgtaccagatgtaattatagagtatttaaatat	408	
Db				
QY 704	GGATCTGACAG	CATATAGTATTGTATTCCTCGTGTGTAATCAGGATTTATTAATAT	763	
Db				
QY 409	gocctgttagaa	tgtttaagacagggcattcagttcggacaaatacgaacccttagt	468	
Db				
QY 764	GATGTAAATAG	GCGCATTAAGAGTAGGGTTATAGACAGTACTAAATACAAACCCATAAG	823	
Db				
QY 469	tatcttgggttc	gaccatgtgcattactcgtattttggcaatgtcctgcctgattattg	528	
Db				
QY 824	ATAGTTGAAT	TAGCTAAATAATACATGTAAATCGTCATGGCGAATTCATGCCCTGACTCTGT	883	
Db				
QY 529	aaaatcagcag	gaggacagaataaaactgtggaattttaagta	571	

	Query Match	15.6%	Score 91.2;	DB 4;	Length 1022;
	Best Local Similarity	53.5%;	Prod. No. 3.3e-21;		
	Matches 214;	Conservative 0;	Mismatches 183;	Indels	Gaps
169	tggcagcttgaagtgtctgtctgcgagagccgagcgagatcgacacatctctgg	228			
529	TGGCAAAATTGAGCTTCACGAGGATTTAATCGCGAACACAGATATCGGAGTATCATCTGG	588			
229	atatgcgagcagcagcagagacgggaagtccgtgtgtgccaaatattctcgactctcaag	288			
590	CTCTATGCTTTACACGAGGAGGAAACACAGCTTCGCGAAGGAATTTTAAATCAGG	645			

[illegible]

Oy	480	cgaccatgtcactcgatttgcgaatgctctgccgtgattatttgaaatacagcag	539
Ddb	875	AGAATATGCGAAGTCATTGTATGCTAACTTCTCCGAGGAAGGAATCTTTTCTGA	934
Oy	540	ggacagaataaaaactg	555
Ddb	935	AGATCGAATAAAGTTG	950

```

RESULT          9
US-08-202-186-9
  : Sequence 9, Application US/08202186
  : Patent No. 5756708
  :
  : GENERAL INFORMATION:
  : APPLICANT: KARAN, Mirko
  : APPLICANT: BURNS, Thomas M.
  : APPLICANT: DALE, James L.
  : APPLICANT: HARDING, Robert M.
  : TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
  : NUMBER OF SEQUENCES: 60
  : CORRESPONDENCE ADDRESS:
  : ADDRESSER: Foley & Lardner
  : STREET: 3000 K Street, N.W.
  : CITY: Washington, D.C.
  : COUNTRY: USA
  : ZIP: 20007-5109
  :
  : COMPUTER READABLE FORM:
  : MEDIUM TYPE: Floppy disk
  : COMPUTER: IBM PC compatible
  : OPERATING SYSTEM: PC-DOS/MS-DOS
  : SOFTWARE: PatentIn Release #1.0, Version #1.25
  :
  : CURRENT APPLICATION DATA:
  : APPLICATION NUMBER: US/08/202,186
  : FILING DATE: 24-FEB-1994
  : CLASSIFICATION: 435
  : ATTORNEY/AGENT INFORMATION:
  : NAME: JEFFERY, Donald D.
  : REGISTRATION NUMBER: 19,980
  : REFERENCE/DOCKET NUMBER: 71611/102 FIVE
  :
  : TELECOMMUNICATION INFORMATION:
  : TELEPHONE: 202 672 5300
  : TELEFAX: 202 672 5399
  : TELEX: 904136
  :
  : INFORMATION FOR SEQ ID NO: 9:
  : SEQUENCE CHARACTERISTICS:
  : LENGTH: 1111 base pairs
  : TYPE: nucleic acid
  : STRANDEDNESS: single
  : TOPOLOGY: circular
  :
  : US-08-202-186-9

```

[illegible]

332	QY	tgaccagtcacatcgaggaccacaaacgaaat	taatactctcgatgaccacagtgtaatt	391
736	DB	GTACATTGTATATTATGAGGATATA	-----GTTATATTGATATCCAGATGCAAG	789
392	QY	tagagttttaaatatgacccctggtcagaat	gtttaagaacacagggcattcagttcggaca	451
790	DB	AGGAATATTTAAACTATGCTCTATTAGA	GAAGAAATTTAAAAATGGAATTTTCAAGCGGGA	849
452	QY	aatacgaaccccttagttatctctgggttcg	accatgctgcatgctactcgtatttgcgaatg	511
850	DB	AATATGACCCGTTTGAANAATTG	---TAGAATATGTGGAAGCATTTGTAATGGCTAACT	906
512	QY	tcctgcctgattatttgaataatcagcaggg	acagacgaataaaact	554
907	DB	TCCTTCGAGGAGGAATCTTTTCTGAAG	ATCGAATCGAATAAAGCT	949

RESULT 13
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-18

	Query Match	10.1%	Score 58.6;	DB 1;	Length 1105;
	Best Local Similarity	52.5%;	Pred. No. 3.7e-10;		
	Matches 180;	Conservative	0; Mismatches 154;	Indels	Gaps
212	atcgcaaatcctctgatatgaggacagacggagacgggaagtccgtgtttgcc	271			
618	ACCGAAGGATTAATTGGGTCTATGGCCCAATCGAGGAAGAAAGACAAC	677			
272	aattctcggactcaagccgcactggttctcacacatggtgtgaaaccagaagga	331			
678	AATAATTTAATGAACGAGAATCCGTTTTTATTCGCCAGGAGAAAATCAT	737			
332	tataccagtatcatcaggaaccacaacgaataattaatcctcgtatgtaccagg	391			

Db	738	GTAGATTGTATTAATTATGAGGAATA-----GTTATATTGATATCCAGATGCAAAAG	791
Qy	392	tagagtatttaaatattgcctgttagaatggttaagaacaggggcatttcagttcggaca	451
Db	792	AGGAATATTAAACATATGTTTTATTAGAGAATTTTAAAGATGGAATTAATCAAAAGCGGGA	851
Qy	452	aataagaaccccttagttatcttgggttcgaccatgtcactcgtatttgcgaatg	511
Db	852	AATATGAACCCGTTTTGAAAATG---TAGAATATGTGGAATCAATTGTATGGCTAACT	908
Qy	512	tcctgcctgatatttgaaaatcagcaggagacagaataaaact	554
Db	909	TCTTTCGGAAGAAGCAATCTTTTCAGAAGATCGCAATAAAGCT	951

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	9.94;	Score 58;	DB 1;	Length 1109;	
	Best Local Similarity	51.14;	Pred. No. 6e-10;			
	Matches 192;	Conservative	0;	Mismatches 175;	Indels	9; Gaps
180	agtgtgtctcgatcgcgagagccagcgagatcgcacaaatctctggtatcgagc	239				
584	AGAGGTGGAGAAATCATATGGCGCACCATGTCATCGGAGATAATTTGGGTCTATGGCCC	643				
240	agacggagggagcgggaagtcggtgtttgcccataatctcggactcaagcccgactggtt	299				
644	AAATGGAGGAGAAGGAAAGACAACTATATGCAAAACGTCTAATGAAGACGAGAAATGCGTT	703				
300	ctacacatgtgtggaaaccagaagaagcagttatgtaccagttacatcaggagaccacaaacg	359				

us-09-462-955-1_copy_409_991.rni

Thu Oct 25 13:08:11 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:22 ; Search time 393.52 Seconds
(without alignments)
607.925 Million cell updates/sec

Title: US-09-462-955-1_copy_611_991

Perfect score: 381

Sequence: 1 cagcgagcagtcgacacatc.....ttaaagtgtgtcatctaa 381

Scoring table:

IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112.4	29.5	1096	AAV71833	BTV DNA II clone
2	112.4	29.5	1096	AAA38950	Banana bunchy top
3	110.8	29.1	1091	AAV71834	BTV DNA II clone
4	110.8	29.1	1091	AAA38947	Banana bunchy top
5	110.6	29.0	886	AAV34687	DNA sequence of BB
6	89.6	23.5	1106	AAV71832	BTV DNA I clone (
7	88.6	23.3	1017	AAT13165	SCSV segment 6. S
8	88.2	23.1	1022	AAT13161	SCSV segment 2. S
9	88	23.1	1106	AAA38946	Banana bunchy top
10	80.4	21.1	593	AAV34686	DNA sequence of BB
11	62.8	16.5	1110	AAV24089	Banana bunchy top

12	59.6	15.6	982	18	AAT49405	Banana bunchy top
13	59.6	15.6	1110	19	AAV24086	Banana bunchy top
14	59.6	15.6	1111	19	AAV24077	Banana bunchy top
15	59.6	15.6	1111	19	AAV24084	Banana bunchy top
16	59.6	15.6	1111	19	AAV24087	Banana bunchy top
17	58.6	15.4	1103	19	AAV24091	Banana bunchy top
18	58.6	15.4	1105	19	AAV24093	Banana bunchy top
19	58	15.2	1109	19	AAV24088	Banana bunchy top
20	58	15.2	1111	19	AAV24090	Banana bunchy top
21	57	15.0	1104	19	AAV24092	Banana bunchy top
22	56.4	14.8	1111	19	AAV24085	Banana bunchy top
23	45.8	12.0	936	22	AAF58252	Oligonucleotide D1
24	45.8	12.0	936	22	AAF58254	Oligonucleotide D1
25	45.8	12.0	936	22	AAF58257	Oligonucleotide D1
26	45.8	12.0	936	22	AAF58259	Oligonucleotide D2
27	45.8	12.0	936	22	AAF58262	Oligonucleotide D2
28	45.8	12.0	938	22	AAF58255	Oligonucleotide D1
29	42.4	11.1	936	22	AAF58252	Oligonucleotide D1
30	42.4	11.1	936	22	AAF58254	Oligonucleotide D1
31	42.4	11.1	936	22	AAF58257	Oligonucleotide D1
32	42.4	11.1	936	22	AAF58259	Oligonucleotide D2
33	42.4	11.1	938	22	AAF58262	Oligonucleotide D2
34	42.4	11.1	938	22	AAF58255	Oligonucleotide D1
35	40.2	10.6	300	20	AAV71831	Subgenomic fragmen
36	40.2	10.6	300	21	AAA38949	Banana bunchy top
37	37	9.7	244	22	AAF58238	Oligonucleotide D1
38	34.2	9.0	244	22	AAF58238	Oligonucleotide D1
39	33.8	8.9	287	20	AAV71830	Subgenomic fragmen
40	33.8	8.9	287	21	AAA38948	Banana bunchy top
41	32.6	8.6	5053	19	AAV22816	Nucleotide sequenc
42	32.6	8.6	5055	19	AAV7236	DNA from a region
43	31.4	8.2	41599	21	AAZ35351	Cosmid including s
44	30.8	8.1	41599	21	AAZ35351	Cosmid including s
45	30.6	8.0	9345	21	AAZ93966	Mouse. uromodulin p

ALIGNMENTS

RESULT 1
AAV71833
ID AAV71833 standard; DNA; 1096 BP.
XX
AC AAV71833;
XX
DT 10-FEB-1999 (first entry)
XX
DE BTV DNA II clone (2-17) nucleotide sequence.
XX
KW Banana bunchy top virus; BTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; ss.
XX
OS Banana bunchy top virus.
XX
FH Key Location/Qualifiers
FT TATA_signal 1..7
FT /*tag= a
FT stem_loop 8..38
FT /*tag= b
FT CDS 70..928
FT /*tag= c
FT /*product= "ORF-V2 product"
FT /*transl_except= (pos:215..217, aa:Gly)
FT polyA_signal 533..538
FT /*tag= d
FT polyA_signal 799..804
FT /*tag= e
FT polyA_signal 907..912
FT /*tag= f
FT polyA_signal 1030..1035
FT /*tag= ge
XX
PN US5846705-A.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:46 ; Search time 5479.82 Seconds
(without alignments)
657.236 Million cell updates/sec

Title: US-09-462-955-1_COPY_611_991
Perfect score: 381
Sequence: 1 cagcgagcagtcgcacaatc.....ttaaagtatgtcatctaa 381

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	40	10.5	512	224	AQ121628	HS_3089_A
2	36.8	9.7	498	231	AQ615572	HS_5144_B
3	36.8	9.7	518	121	AW818171	CMT-ST027
4	35.6	9.3	318	224	AQ089903	HS_3001_A
5	35.4	9.3	460	104	AI942816	fc67b05.x
6	35.4	9.3	955	137	BE603735	HVSMH008
7	35.2	9.2	345	243	AZ401897	IM0168N22
8	35.2	9.2	848	233	AQ746046	HS_2277_A
9	34.6	9.1	487	228	AQ418525	RPCI-11-2
10	34.6	9.1	555	228	AQ417598	RPCI-11-2
11	34.6	9.1	1101	219	CNS016H0	RPCI-11-2
12	34	8.9	443	223	AQ010299	HS_3249_A
13	33.8	8.9	410	233	AQ0807176	HS_3249_A
14	33.8	8.9	604	219	AQ019820	Homo sapi
15	33.8	8.9	772	229	AQ488982	RPCI-11-2
16	33.8	8.9	894	221	CNS032CO	RPCI-11-2
17	33.6	8.8	326	223	AQ008775	Tetraodon
18	33.6	8.8	418	231	AQ632834	RPCI-11-2
19	33.6	8.8	605	229	AQ481591	RPCI-11-2
20	33.6	8.8	658	224	AQ077737	CIT-HSP-2
21	33.6	8.8	749	241	AZ331069	IM0056P10
22	33.4	8.8	348	227	AQ348508	RPCI-11-1
23	33.4	8.8	411	17	A1226907	ujl1b01.y
24	33.4	8.8	517	230	AQ565229	HS_5355_B
25	33.4	8.8	532	230	AQ565203	HS_5355_B
26	33.4	8.8	637	227	AQ377093	RPCI-11-14
27	33.4	8.8	751	233	AQ745275	HS_5503_A
28	33.4	8.8	764	226	AQ240040	CIT-HSP-2
29	33.4	8.8	785	192	AK002866	Mus muscu
30	33.2	8.7	459	239	AZ141666	SP-0014_B
31	33.2	8.7	617	137	BE585743	Est#1SP6
32	33.2	8.7	690	137	BE585745	Est#1SP6
33	33.2	8.7	840	235	AQ896731	HS_3134_A
34	33	8.7	503	233	AQ800876	HS_5258_B
35	33	8.7	813	239	AZ190921	SP-1018_B
36	32.8	8.6	224	159	N49654	YV26B10.11
37	32.8	8.6	471	225	AQ229338	HS_2022_B
38	32.6	8.6	502	225	AQ209803	HS_3244_A
39	32.6	8.6	525	225	AQ189601	HS_3219_B
40	32.6	8.6	569	173	BG052823	RHI22.14
41	32.6	8.6	614	230	AQ590659	HS_5398_A
42	32.6	8.6	939	220	CNS02TRL	Tetraodon
43	32.4	8.5	415	224	AQ087707	HS_2234_B
44	32.4	8.5	425	150	BF544163	UI-R-EO-d
45	32.4	8.5	441	225	AQ234990	HS_2056_A

ALIGNMENTS

RESULT 1					
AQ121628/c					
LOCUS	AQ121628	512 bp	DNA	22-SEP-1998	
DEFINITION	HS_3089_AL_B03_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3089 Col=5 Row=C, DNA sequence.				
ACCESSION	AQ121628				
VERSION	AQ121628.1	GI:3498794			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
AUTHORS	1 (bases 1 to 512)				
	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellner,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and				

JOURNAL MEDLINE COMMENT

Scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3089 row: C column: 5

Class: BAC ends

High quality sequence stop: 512.

Location/Qualifiers

1..512

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="plate=3089 Col=5 Row=C"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 174 a 85 c 115 g 136 t 2 others

ORIGIN

Query Match 10.5%; Score 40; DB 224; Length 512;

Best Local Similarity 59.8%; Pred. No. 0.14;

Matches 67; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 268 atcttggttcgacacgtgcatgtactgtatctgctgcaatgctgctgctgattttgga 327

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 179 ATCTTATTGCGCCATCTGCTTTACTCCACTTATACCAATGCTTTTAAAGTATCCCA 120

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 328 aaatcagcagggacagacataaaactgtggaatatttaagtgatgtgcatct 379

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 119 AACTGAACATAGATATGATCAGACCTCAGGAAGATTAAATGAACACACATTT 68

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2

AQ615572

LOCUS

DEFINITION

HS_5144_B1_B02_SP6E RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plate=720 Col=3 Row=D, DNA sequence.

ACCESSION

AQ615572

VERSION

AQ615572.1 GI:5076848

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 498)

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellner,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

High Throughput Sequencing Center

University of Washington

Contact: Mahairas GG, Wallace JC, Hood L

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

http://www.htsc.washington.edu

BASE COUNT	193 a	119 c	89 g	117 t
ORIGIN				

	Query Match	9.7%	Score 36.8	DB 121	Length 518
	Best Local Similarity	61.5%	Pred. No. 1.3		
	Matches 59	Conservative 0	Mismatches 0	Indels 37	Gaps 0
QY	145	aggaccacaaacgaatttaactctctgtagtaccacaggtgttaatttagaggtatttaatt	204		
Db	168	AGGATTAACAAACCAATTTTGATCTCATCTGGACACAAACAAATATATAGTAACAAGATT	227		
QY	205	atgcctgttagaagtgttaagaacagggcattca	240		
b	238	ccccctacacacaaacgtctttaaagcctgcccccatca	263		

RESULT	4
AQ089903/C	
LOCUS	318 bp
DEFINITION	DNA
	GSS
	26-AUG-1998
ACCESSION	AQ089903
VERSION	HS_3001_A2_H04_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3001 Col=8 Row=O, DNA sequence.
KEYWORDS	AQ089903
SOURCE	AQ089903.1 GI:3458814 GSS. human.

ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 318)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: iwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3001 row: 0 column: 8
Class: BAC ends
High quality sequence stop: 318.

FEATURES	source	Location/Qualifiers
1..318		
/organism="Homo sapiens"		
/db_xref="taxon.9606"		
/clone="plate=3001 Col=8 Row=O"		
/clone_lib="CIT Approved Human Genomic Sperm Library D"		
/sex="male"		
/note="Organ: sperm; Vector: pBeloBAC11; BAC clones in E-Coli DH10B"		
77 a	66 g	108 t
BASE COUNT	66 C	1 others

Query Match 9.3%; Score 35.6; DB 224; Length 318;
Best Local Similarity 54.6%; Pred. No. 2.8;

```

Plate: 720 row: D column: 3
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 498.
Location/Qualifiers
1. .498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=720 Col=3 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"

```

BASE COUNT	171 a	107 c	87 g	113 t	20 others
ORIGIN					

Query Match 9.7%; Score 36.8; DB 231; Length 498;
Best Local Similarity 56.7%; Pred. No. 1.3;
Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

272 gtraagaacagggcattcagttcgagacaatacgaacccttagttatcttgggttcgac 281
273 cccccccccccgagcaatcagccgacacaaagaataaaggcgatttaagttaggaaaaagag 330

[illegible]

Db 331 GAAGTCAAATTGTCCTGTTGCAGATGACATGATGTTTATTATAGAAATCCCCCATCGAC 390

RESULTS	AW18171	LOCUS	518 bp	mRNA	EST	17-MAY-2000
DEFINITION	CM1-ST0277-161299-070-h05	ST0277	Homo sapiens	cdna, mRNA sequence.		

ACCESSION AW818171
VERSION AW818171.1 GI:7911165
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 518)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S. Simpson, D. H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
50, Victoria Road, 100 Victoria Park, Singapore 01500-010

Rua Prof. Antonio Prudente 109, 4 andar, 01309-010, São Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
<http://www.ludwig.cancer.br/genetics/genotkml2.pl?l21=162?CM1=SM02777-161>

299-070-h05&t3=1999-12-16&t4=1)
 (http://www.luowig.org.br/scripts/getcunhaiz.pl?cr=cz-cm1-310277-10)
 seq primer: puc 18 forward
 High quality sequence start: 76

FEATURES	source
High quality sequence stop: 183.	
Location/Qualifiers	
1..518	
organism="Homo sapiens"	

	Matches	71;	Conservative	0;	Mismatches	59;	Indels	0;	Gaps	0;
Qy	222	gttaagaacagcggcattcagttcgcgacaaatcagaacccctagttatcttgggttcgac	281							
Db	212	gTTCtGCCcAGGcCAATcAGcAGcAGcAGAAATAAGGGTATTTAATTAGcAGcAGAG	153							
Qy	282	catgtcgtactcgtatttgcgcaatgcctgcctgattatttgaaataatcagcagggac	341							
Db	152	GATTTCAAATTGTCCTGTTTCAGATGCATGACTGTATATCTAGAAAAACCCGCAATTGTC	93							
Qy	342	agaataaaac	351							
Db	92	TCAGTCCCAAC	83							

RESULT	5
AI942816	
LOCUS	.. 460 bp mRNA
DEFINITION	fc57b05.x1 zebrafish wasnu MPIMG EST Danio rerio cDNA 3' , mRNA
ACCESSION	AI942816
VERSION	AI942816.1 GI:5707472
KEYWORDS	EST.
SOURCE	zebrafish.
ORGANISM	Danio rerio
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasborinae; Danio.

embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

	CONC'L.		
BASE COUNT	154 a	60 c	59 g
ORIGIN			187 t

Query Match	9.3%	Score 35.4;	DB 104;	Length 460;
Best Local Similarity	50.9%	Pred. NO. 3.5;		
Matches 84;	Conservative	0;	Mismatches	91;

RESULT 6

BE603735	LOCUS	BE603735	955 bp	mRNA	EST	02-MAR-2001
	DEFINITION	HVSMEN0087003f	Hordeum vulgare	5-45 DAP spike	EST library	
		HVCDA0009 (5 to 45 DAP)	Hordeum vulgare	cDNA clone	HVSMEN0087003f,	
		mRNA sequence				

BASE COUNT
ORIGIN

BASE COUNT	228 a	136 c	117 g	144 t	223 others
ORIGIN					
Query Match		9.2%	Score 35.2;	DB 233;	Length 848;
Best Local Similarity		55.8%;	Pred. No. 4.5;		
Matches	67;	Conservative	0;	Mismatches 53;	Indels 0; Gaps 0

was hydrodynamically sheared by repeated passages through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative


```

QY 222 gtaagaacaggcattcagtcgacaaatagcaaccccttagttatcttggttcac 281
    ||| | ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 334 GTTCTGGCCAGGCAATCAGTAGGAGAAATAAACCAGTATTCATATAGGAAAGAG 393
    ||| | ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 282 catgtgcattctcgtatttgcaatgctcctgctgatttgaataacagcagggac 341
    ||| | ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 394 GAAGTCAAAATGTCCCTGTTTGCAGATGACATGATTGTATATTAGAAAACCCCATTTGC 453

RESULT 9
AQ418525 AQ418525 487 bp DNA GSS 23-MAR-1999
LOCUS RPCI-11-20318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-20318,
DEFINITION DNA sequence.
ACCESSION AQ418525
VERSION AQ418525.1 GI:4476249
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 487)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Map Building
COMMENT Other_GSSs: RPCI-11-20318.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES
Source Location/Qualifiers
1..487
/organism="Homo sapiens"
/db_xref="GDB:757767"
/db_xref="taxon:9606"
/clone="RPCI-11-20318"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 165 a 102 c 96 g 124 t
ORIGIN

Query Match 9.1%; Score 34.6; DB 228; Length 487;
Best Local Similarity 54.3%; Pred. No. 6.2;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 208 cccctttagaatgtgttaagaacaggcattcagtcgacaaatagcaaccccttagtt 267
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
Db 127 CTCTGGTAGAATTCGTCCTCCGTCAGGCAATCAGGAGGAGAAAGAAATAAAGGCTTC 186
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
QY 268 atcttgggttcacattgcatgtactcgtatttgccaatgctcctgctgattattga 327
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
Db 187 AATTAGGAAAAGAGGAAGTCAAAATGTGCCCTGTTTGCAGATGACATGATTGTGTATCTAC 246
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
QY 328 aaatcagca 336
    ||| | ||

```

```

Db 247 AAAACCCCA 255

RESULT 10
AQ417598 AQ417598 555 bp DNA GSS 23-MAR-1999
LOCUS RPCI-11-203C8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203C8,
DEFINITION DNA sequence.
ACCESSION AQ417598
VERSION AQ417598.1 GI:4475441
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 555)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Map Building
COMMENT Other_GSSs: RPCI-11-203C8.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES
Source Location/Qualifiers
1..555
/organism="Homo sapiens"
/db_xref="GDB:757762"
/db_xref="taxon:9606"
/clone="RPCI-11-203C8"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 192 a 114 c 110 g 139 t
ORIGIN

Query Match 9.1%; Score 34.6; DB 228; Length 555;
Best Local Similarity 54.3%; Pred. No. 6.3;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 208 cccctttagaatgtgttaagaacaggcattcagtcgacaaatagcaaccccttagtt 267
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
Db 150 CTCTGGTAGAATTCGTCCTCCGTCAGGCAATCAGGAGGAGAAAGAAATAAAGGCTTC 209
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
QY 268 atcttgggttcacattgcatgtactcgtatttgccaatgctcctgctgattattga 327
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
Db 210 AATTAGGAAAAGAGGAAGTCAAAATGTGCCCTGTTTGCAGATGACATGATTGTGTATCTAC 269
    ||| ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| |
QY 328 aaatcagca 336
    ||| | ||
Db 270 AAAACCCCA 278

RESULT 11
CNS016H0 CNS016H0 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION

```

[illegible]

Db	311	CTAATAAACANTTCAAGTCGACTCTTAACACAGGTGAATTTATAGTGCCAAATTATATA	370
Qy	210	ctgttagaatgtggttaagaacagg	234
Db	371	CTAATAAAAACCTGTTTTAAAAAAGG	395
RESULT	15		
AQ488982		772 bp	DNA
LOCUS		RPCI-11-267P13.TJ	RPCI-11 Homo sapiens genomic clone RPCI-11-267P13
DEFINITION		DNA sequence.	
ACCESSION		AQ488982	GSS
VERSION		AQ488982.1	GI:4674856
KEYWORDS		GSS.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.	
TITLE		Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building	
JOURNAL		Unpublished (1997)	
COMMENT		Other_GSSs: RPCI-11-267P13.TV Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbe@tigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jeong.med.buffalo.edu). Clones may be purchased from PACAPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html . Seq primer: SP6 Class: BAC ends.	
FEATURES		Location/Qualifiers	
SOURCE		1..772	
		/organism="Homo sapiens"	
		/db_xref="GDB:7602516"	
		/db_xref="taxon:9606"	
		/clone="RPCI-11-267P13"	
		/clone_lib="RPCI-11"	
		/sex="Male"	
		/cell_type="Lymphocytes"	
		/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"	
BASE COUNT		300 a 172 c 122 g	178 t
ORIGIN			
Query Match		8.9%;	Score 33.8; DB 229; Length 772;
Best Local Similarity		53.4%;	Pred. No.12;
Matches	71;	Conservative	0; Mismatches 62; Indels 0; Gaps 0;
QY	222	gttaaacacagggcatttcagttcggaacaatcacgaacccttagttatcttggttcgcac	281
Db	514	GTTCGTGCAGGGCAATCAGCGAAGAAGAAATAACGGATATTCAGTTACGAAAAGAG	573
QY	282	cattgtcactgtactcgtatttgccaaatgctcctgcctgatatttgaagcaagcaggac	341
Db	574	GAAGTCAAAATGTCTCTGTTTCGATCATGACTGATATTTAGAAAATCCCATCGTC	633
QY	342	aqaataaaaactgt	354
Db	634	TCAGCCCAATCT	646

us-09-462-955-1_copy_611_991.rst

Thu Oct 25 13:08:25 2001

Search completed: October 24, 2001, 13:13:47
Job time: 11812 sec

. . . .

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:26 ; Search time 180.6 Seconds
(without alignments)
399.377 Million cell updates/sec

Title: US-09-462-955-1_COPY_611_991

Perfect score: 381
Sequence: 1 cagcgacgacgacgacacac.....ttaaagtatgtcatctaa 381

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112.4	29.5	1096	2	US-08-418-071-4
2	110.8	29.1	1091	2	US-08-418-071-5
3	89.6	23.5	1106	2	US-08-418-071-3
4	88.6	23.3	1017	4	US-08-793-634B-6
5	88.2	23.1	1022	4	US-08-793-634B-2
6	62.8	16.5	1110	1	US-08-202-186-14
7	59.6	15.6	982	3	US-08-973-068-28
8	59.6	15.6	1111	1	US-08-202-186-11
9	59.6	15.6	1111	1	US-08-202-186-9
10	59.6	15.6	1111	1	US-08-202-186-12
11	59.6	15.6	1111	1	US-08-202-186-24
12	58.6	15.4	1103	1	US-08-202-186-16
13	58.6	15.4	1105	1	US-08-202-186-13
14	58	15.2	1109	1	US-08-202-186-18
15	58	15.2	1111	1	US-08-202-186-15
16	57	15.0	1104	1	US-08-202-186-17
17	56.4	14.8	1111	1	US-08-202-186-10
18	40.2	10.6	300	2	US-08-418-071-2
19	33.8	8.9	287	2	US-08-418-071-1
20	29.4	7.7	8920	4	US-08-446-855A-1
21	29.4	7.7	8920	4	US-09-150-741-1
22	28.8	7.6	10607	1	US-08-078-090-3
23	28.4	7.5	1380	4	US-09-499-505-4
24	28.4	7.5	1430	4	US-09-499-505-3
25	28	7.3	28958	1	US-08-258-261B-6
26	28	7.3	28958	1	US-08-456-837-6
27	28	7.3	28958	1	US-08-457-342-6

28 28 7.3 28958 1 US-08-457-646A-6 Sequence 6, Appli
29 28 7.3 28958 1 US-08-458-076A-6 Sequence 6, Appli
30 28 7.3 28958 1 US-08-764-233A-4 Sequence 4, Appli
31 28 7.3 28958 1 US-08-457-335A-6 Sequence 6, Appli
32 28 7.3 28958 1 US-08-729-214-6 Sequence 6, Appli
33 28 7.3 28958 3 US-09-028-934-6 Sequence 6, Appli
34 28 7.3 49377 1 US-08-764-233A-1 Sequence 1, Appli
35 27.8 7.3 2570 2 US-09-056-075-2 Sequence 2, Appli
36 27.6 7.2 2899 2 US-08-624-581-3 Sequence 2, Appli
37 27.6 7.2 2917 2 US-08-624-581-3 Sequence 3, Appli
38 27.6 7.2 2959 2 US-08-624-581-1 Sequence 1, Appli
39 27.6 7.2 6476 4 PCW-US91-08177-12 Sequence 5, Appli
40 27.4 7.2 840 5 PCW-US91-08177-12 Sequence 12, Appli
41 27.4 7.2 2256 6 5220013-1 Patent No. 5220013
42 27.4 7.2 7323 5 PCT-US91-08177-1 Sequence 1, Appli
43 27.2 7.1 5892 3 US-08-755-587-27 Sequence 27, Appli
44 27.2 7.1 7240 3 US-08-755-587-15 Sequence 15, Appli
45 27.2 7.1 11283 2 US-08-603-753D-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-418-071-4
; Sequence 4, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418.071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-4

; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: DCB-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 596-9000
 ; TELEFAX: (212) 596-9090
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1106 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: genomic DNA
 ; DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 ; US-08-418-071-3

Query Match 23.5%; Score 89.6; DB 2; Length 1106;
 Best Local Similarity 52.7%; Pred. No. 3.8e-21;
 Matches 194; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

```

QY 2 agcggacgatgcacaaatcctctggatcgacgacgagacggagacggagagtcctcg 61
Db 559 ACCGGATGATCGAAGCATCATCTGGGTCTATGGTCTTATGGTAAATCAGGGTAAATCAAC 618
QY 62 gttgcacaaatcctcgactcaagcccgactgttctcacatgtggtggaaaccagaaa 121
Db 619 ATATGCCAAGTCACTAATCAAGAAGGATGTTGTTCTACACCGGGTGGGAAGAGAGAA 678
QY 122 ggaactgttaccagtcacatcgaggaaccccaacgaaatttaattcctcgatgtacc 181
Db 679 TATCTTATTCCTCTATGTGACCAAGGATCTGACACATATAGTATTTCATATTCCTCG 738
QY 182 gtgaatttagagatttaaatattccctgttgcgaatgtgttaagaacagggttcag 241
Db 739 TTGTAATCAGGATTTTAAATTTATGTAATAGAGGCATTAAAGGATAGGTTATAGA 798
QY 242 ttcggacaaatacgaaccccttagttatcttgggttcgaccatgtcactgtactatt 301
Db 799 GAGTACTAAATACAAACCCATAAGATAGTTGAAATAGGTAAATACATGTAATCGTAT 858
QY 302 tgcaatgtcctcctgattatttgaataacgacgagcagacagaataaaaactgtggaat 361
Db 859 GCGCAATTTTCATGCTGCTGACTCTGTAAATCTCCGAAGATCGAATAAAATCATTTATTG 918
QY 362 ttaaaata 369
Db 919 CTGAAGAA 926
  
```

RESULT 4

US-08-793-634B-6
 ; Sequence 6, Application US/08793634B
 ; Patent No. 6211431

; GENERAL INFORMATION:
 ; APPLICANT: Boevink, Petra C.
 ; APPLICANT: Surin, Brian P.
 ; APPLICANT: Keese, Paul K.
 ; APPLICANT: Chu, Paul W.G.
 ; APPLICANT: Waterhouse, Peter M.
 ; APPLICANT: Khan, Rafiqul I.
 ; APPLICANT: Larkin, Philip J.
 ; APPLICANT: Taylor, William C.
 ; APPLICANT: Marchall, Jerry S.
 ; TITLE OF INVENTION: NOVEL PLANT PROMOTERS AND USES
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City

; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/793,634B
 ; FILING DATE: June 9, 1997
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digiglio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 10530
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 516-742-4343
 ; TELEFAX: 516-742-4366
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1017 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-793-634B-6

Query Match 23.3%; Score 88.6; DB 4; Length 1017;
 Best Local Similarity 54.1%; Pred. No. 8e-21;
 Matches 203; Conservative 0; Mismatches 169; Indels 3; Gaps 1;

```

QY 2 agcggacgatgcacaaatcctctggatcgacgacgagacggagacggagagtcctcg 61
Db 539 ACCAGATTATAGACGATAAATCTGGTCTATGGACCTCTGCTGTAATGAAGCAATCTAC 598
QY 62 gttgcacaaatcctcgactcaagcccgactgttctcacatgtggtggaaaccagaaa 121
Db 599 ATTTGCAACAGCATCTGTCTATTGAAAGATGGTTGGGGTTTATCTGCTGGAGGAAGACACA 658
QY 122 ggaactgttaccagtcacatcgagaccccaacgaaatttaattcctcgatgtacc 181
Db 659 AGATATGATGTCATCTGTGACTGCTGACCTTAAGAAATTAATTTGGGTATTGACATACCCAG 718
QY 182 gtaatttagagattttaaattatgcctgttgcgaatgtgttaagaacagggttcag 241
Db 719 AGTTAGTTTCAAGATGATGTGAATTTATGGTGTAAATAGAACAGGTTAAGAATAGGGTAATGGT 778
QY 242 ttcggacaaatacgaaccccttagttatcttgggttcgaccatgtcactgtactcgt 298
Db 779 GAATACTAAGTATGAGCCATGTGTAATGCGGGATGATAATCATCTCTGTTTATGTAATGT 838
QY 299 atttgcaaatgtcctgcctgattatttgaataacgacgagcagacagaataaaaactgtgaa 358
Db 839 GTTTGCAAAATGCTACTCCAGATTTGGGAAATTAAGTCAAGATACATAAATAAATTAATTCG 898
QY 359 tatttaaaagtatgtg 373
Db 899 TTGTTGAAAACTCTG 913
  
```

RESULT 5

US-08-793-634B-2
 ; Sequence 2, Application US/08793634B
 ; Patent No. 6211431
 ; GENERAL INFORMATION:
 ; APPLICANT: Boevink, Petra C.
 ; APPLICANT: Surin, Brian P.
 ; APPLICANT: Keese, Paul K.
 ; APPLICANT: Chu, Paul W.G.
 ; APPLICANT: Waterhouse, Peter M.
 ; APPLICANT: Khan, Rafiqul I.

Best Local Similarity	53.2%	Pred. NO. 3.5e-12;	2;
Matches	183.	Conservative	9: Gaps
		0: Mismatches	Indels
		152:	Indels

2 agcggacgatacgcacaatcctctggatatcgggacgagacggaggagacgggaagtcct 61

Best Local Similarity 54.5%; Pred. No. 1.1e-20;
Matches 100. Conservative 0: Mismatches 163: Indels 1:
Gaps 1:

US-08-202-186-14


```

: Patent No. 6127604
:
: GENERAL INFORMATION:
:
: APPLICANT: Dale, James Langham
: APPLICANT: Harding, Robert Maxwell
: APPLICANT: Dugdale, Benjamin
: APPLICANT: Beetham, Peter Ronald
: APPLICANT: Hafner, Gregory John
: APPLICANT: Becker, Douglas Kenneth
:
: TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
:
: FILE REFERENCE: 09657/002001
:
: CURRENT APPLICATION NUMBER: US/08/973,068
: CURRENT FILING DATE: 1998-03-12
: EARLIER APPLICATION NUMBER: PCT/AU96/00335
: EARLIER FILING DATE: 1996-05-31
:
: NUMBER OF SEQ ID NOS: 61
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ ID NO 28
:
: LENGTH: 982
:
: TYPE: DNA
:
: ORGANISM: Banana Bunchy Top Virus (BBTV)
:
: US-08-973-068-28

```

	Query Match	15.6%	Score 59.6;	DB 3;	Length 982;
	Best Local Similarity	52.6%;	Pred. No. 6.4e-11;		
	Matches 181;	Conservative . 0;	Mismatches 154;	Indels	Gaps 2;
QY	10	atcgcacaatctcttgatatcgagcagacgaggagagacgggaagtcgtgtttgccca	69		
Db	85	atcggagaataatttggctctatgcccgaatggaggagaagaaacacagtatgcaa	144		
QY	70	aatatctcggactcaaacccgactggtctcacacatgtggtggaaccagaagagcgcat	129		
Db	145	aacatctaataagacgagaaatcgcttttatctccaggaggaataatcattggatatat	204		
QY	130	tgtaccagtaacatcgagagccccaaaacgaatttaactctcatgtaccacagtgtaatt	189		
Db	205	gtagactgtaaatcacaggatat-----tgttatattgatattccaagatgcagaag	258		
QY	190	tagagtatttaaattatgccttttagaatgtgtttaagAACAGGGCATTCTCGGACA	249		
Db	259	aggactattaaattatgggtttattgaggaatttaagaatggaataattcaaacggga	318		
QY	250	aatacgaacctctagttatcttgggttcgaccatgtgatctactgattttgccaatg	309		
Db	319	aatatgaacccgttttgaagatag---tagaatgtcgaagccattttaatggctaa	375		
QY	310	tcttgctctgattatttgaataatcagcaggagacagaataaaaaactg	353		
Db	376	tccftccgaaggagaatactttttctgaagatcgaataaaatttg	419		

RESULT 8
 US-08-202-186-11
 : Sequence 11, Application US/08202186
 : Patent No. 5756708
 : GENERAL INFORMATION:
 : APPLICANT: KARAN, Mirkio
 : APPLICANT: BUENS, Thomas M.
 : APPLICANT: DALE, James L.
 : APPLICANT: HARDING, Robert M.
 : TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
 : NUMBER OF SEQUENCES: 60
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W.
 : CITY: Washington, D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/202,186
 ; FILING DATE: 24-FEB-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: JEFFERY, Donald D.
 ; REGISTRATION NUMBER: 19,980
 ; REFERENCE/DOCKET NUMBER: 71611/102 FKE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202 672 5300
 ; TELEFAX: 202 672 5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1110 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ;
 ; US-08-202-186-11

Query Match	15.6%	Score	59.6	DB J	Length	1110			
Best Local Similarity	52.6%	Pred.	No. 6	8e-11					
Matches	181	Conservative	0	Mismatches	154	Indels	9	Gaps	
QY	10	atcgacaaatcctctgatatgctgcgacgagcaggagcagggaatccctgtgttgcca	69						
Db	616	ATPCGGAGAATAATTTGGGTCTATGGACCAATATGGAGGAGAGGAAGACACACCTATGCCAA	675						
QY	70	aatatctcggactcaagcccgactgcttctacacatgctggttgaaacccagaaagacglat	129						
Db	676	RACATCTAATGACACGAGAAATTCGCTTTTATTCCTCCAGGAGGAAAAATCATGCGATATAT	735						
QY	130	tgtaccagttacatcgagaccccaaacgaatttaactcctcgatgtaccacagtgtaatt	189						
Db	736	GTAGACTGTATAATTACGAGGATAT-----TGTTATATTTGATATTTCCAAGATGCAAG	789						
QY	190	tagagtatttaattatgccctgttagaatgtgttaagacacagggcattcagttcgcca	249						
Db	790	AGGATTTATTAATTTATGGTATTATAGAGGAATTTTAGAATGGAATTAATTTCAAGCGCGGA	849						
QY	250	aatcgaaccccttagtctaacttggttcgaccatgtgcgtactcgtatttggccaatg	309						
Db	850	AATATGAACCCGTTTCAAGATAG--TAGAATATGTCGAAGTCATTGTAAATGGCTAACT	906						
QY	310	tctgcctgattattgaaatacagcagggacagaataaaactg	353						
Db	907	TTCTTCCGAAGGAAGAAATCTTTTCTGAAGATCGAAATAAGTTG	950						

```

RESULT 9
US-08-202-186-9
; Sequence 9, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1 25

```

Query Match 15.6%; Score 59.6; DB 1; Length 1111;
Best Local Similarity 52.6%; Pred. No. 6.8e-11;
Matches 181; Conservative 0; Mismatches 154; Indels 9

```

RESULT 11
US-08-202-186-24
; Sequence 24, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186

```

Query Match	15.6%;	Score 59.6;	DB 1;	Length 1111;
Best Local Similarity	52.6%;	Pred. No. 6.8e-11;		
Mismatches 181. Conservative	0;	Mismatches 154;	Indels 9;	Gaps 2;

```

RESULT 10
US-08-202-186-12
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```



```

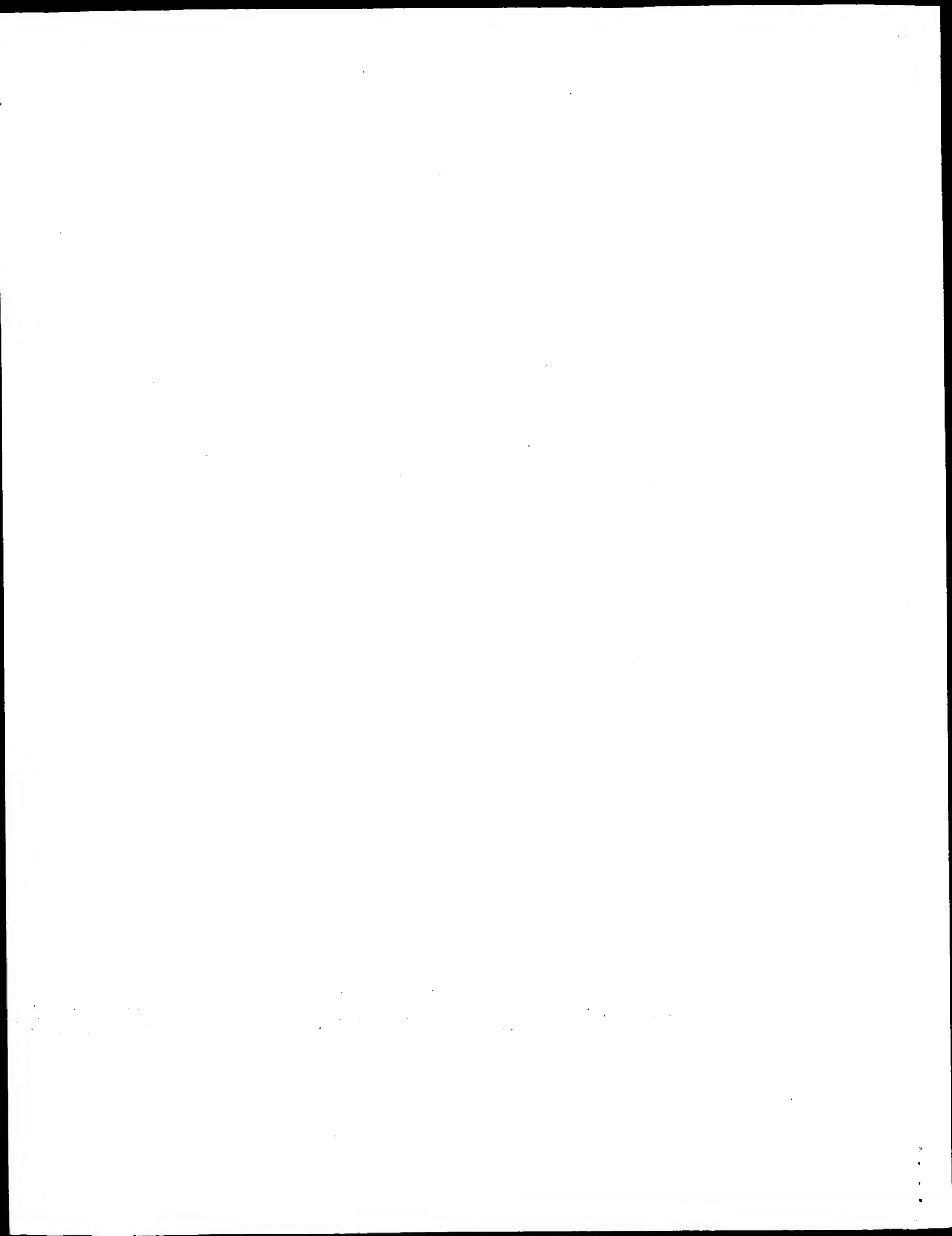
RESULT 15
US-08-20186-15
; Sequence 15, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994

```

```

RESULT 14
US-08-202-186-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
;
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202.186

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:24 ; Search time 393.52 Seconds
(without alignments)
448.364 Million cell updates/sec

Title: US-09-462-955-1_COPY_711_991

Perfect score: 281

Sequence: 1 cacatgtgtggaacagaa.....ttaaagtatgtcatctaa 281

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601:*

```

1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.8	25.2	1091	20	AAV71834
2	70.8	25.2	1091	21	AAA38947
3	70.8	25.2	1096	20	AAV71833
4	70.8	25.2	1096	21	AAA38950
5	70.8	25.2	1106	20	AAV71832
6	70.6	25.1	886	20	AAA34687
7	69.2	24.6	1106	21	AAA38946
8	67.4	24.0	593	20	AAA34686
9	65.2	23.2	1017	17	AAT13165
10	61.8	22.0	1022	17	AAT13161
11	44.6	15.9	1105	19	AAV24093

12	44.4	15.8	1103	19	AAV24091	Banana bunchy top
13	43.8	15.6	936	22	AAF58252	Oligonucleotide D1
14	43.8	15.6	936	22	AAF58254	Oligonucleotide D1
15	43.8	15.6	936	22	AAF58257	Oligonucleotide D1
16	43.8	15.6	936	22	AAF58259	Oligonucleotide D2
17	43.8	15.6	936	22	AAF58262	Oligonucleotide D1
18	43.8	15.6	938	22	AAF58255	Oligonucleotide D1
19	43	15.3	1110	19	AAV24089	Banana bunchy top
20	43	15.3	1111	19	AAV24087	Banana bunchy top
21	42.8	15.2	1104	19	AAV24092	Banana bunchy top
22	41.8	14.9	936	22	AAF58252	Oligonucleotide D1
23	41.8	14.9	936	22	AAF58254	Oligonucleotide D1
24	41.8	14.9	936	22	AAF58257	Oligonucleotide D1
25	41.8	14.9	936	22	AAF58259	Oligonucleotide D2
26	41.8	14.9	936	22	AAF58262	Oligonucleotide D2
27	41.8	14.9	938	22	AAF58255	Oligonucleotide D1
28	41.4	14.7	982	18	AAV49405	Banana bunchy top
29	41.4	14.7	1109	19	AAV24088	Banana bunchy top
30	41.4	14.7	1111	19	AAV24077	Banana bunchy top
31	41.4	14.7	1111	19	AAV24084	Banana bunchy top
32	40	14.2	1111	19	AAV24085	Banana bunchy top
33	39.8	14.2	1110	19	AAV24086	Banana bunchy top
34	39.8	14.2	1111	19	AAV24090	Banana bunchy top
35	37	13.2	244	22	AAF58238	Oligonucleotide D1
36	34.2	12.2	244	22	AAF58238	Oligonucleotide D1
37	32.6	11.6	5053	19	AAV22816	Nucleotide sequence
38	32.6	11.6	5055	19	AAV17236	DNA from a region
39	31.4	11.2	41599	21	AAZ35351	Cosmid including s
40	30.8	11.0	41599	21	AAZ35351	Cosmid including s
41	30.6	10.9	9345	21	AAZ93966	Mouse uromodulin p
42	29.8	10.6	1510	17	AAV29628	DNA mismatch repai
43	29.6	10.5	811	21	AAC50650	Arabidopsis thalia
44	29.4	10.5	8920	15	AAQ62924	Carbamoyl-phosphat
45	29.2	10.4	148	21	AAA41338	Human secreted exp

ALIGNMENTS

RESULT 1

AAV71834
ID AAV71834 standard; DNA; 1091 BP.
XX
AC AAV71834;
XX
10-FEB-1999 (first entry)
XX
BTVD DNA II clone (2) nucleotide sequence.
DE
BTVD DNA II clone (2) nucleotide sequence.
XX

Banana bunchy top virus; BBTVD; PCR technique; plant; tissue; banana;
Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
OS
Banana bunchy top virus.
XX
US5846705-A.
XX
08-DEC-1998.
XX
06-APR-1995; 95US-0418071.
XX
06-APR-1995; 95US-0418071.
XX
(BIOT-) DEV CENT BIOTECHNOLOGY.
XX
Soong T, Wu R, You L;
XX
WPI; 1999-059037/05.
XX

Nucleic acids having banana bunchy top virus component sequences -
used to design primers for use in polymerase chain reaction
detection of the virus
XX
Disclosure; Fig 12A-C; 27pp; English.
PS

XX This represents the nucleotide sequence of a banana bunchy top virus
CC (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention
CC provides nucleic acid sequences associated with BBTV that can be used in
CC a PCR technique for detecting BBTV. The nucleic acid sequences (AAV71830
CC to AAV71833) are used as the basis for the construction of PCR primers,
CC to detect BBTV infection. The PCR technique is used for detecting BBTV
CC in plant tissues (preferably banana, especially Musa species). The
CC virus, one of the most important banana species viruses, causes phloem
CC damage and is transmitted by aphids. PCR detection gives accurate,
CC reliable and specific determination of absence or presence of the
CC virus.
XX Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T; 0 other;
SQ

Query Match 25.2%; Score 70.8; DB 20; Length 1091;
Best Local Similarity 56.2%; Pred. No. 4.1e-13;
Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;
XX

QY 8 ggtggaaccagaaagcgtattgtaccagtcacatcgaggaccacaaacgaaatttaac 67
DB 663 ggtggaagcgtcggtatgatgcacatcaaacgagtgatcctgataatcattggatt 722
QY 68 ctccggtaccaggtgtaatttagagttattaaattatgcctgttagaattgtttaag 127
DB 723 attgataccccagagtcattcagattatctgaattatggtggttatagacaaattaag 782
QY 128 acagggcattcagttcgacaaatacgaacc---ccttagttatcttgggttcaccat 184
DB 783 aatagagtttttaataataacaaatacgaaccatggtgattagaaaagatggacaaaat 842
QY 185 gtccatgtactcgtatttgcgaatgtcctgcctgattatttgaatacagcaggacaga 244
DB 843 gtccatgtattgttaggcaaatgtgtgctgattattgttaaaatttcagaagataga 902
QY 245 ataaactgtggaattatttaagtagtattgtgtcattc 278
DB 903 ataaaaataattattgttgagaaggaacttc 936

RESULT 2
AAA38947
ID AAA38947 standard; DNA; 1091 BP.
XX
AC AAA38947;
XX
DT 23-AUG-2000 (first entry)
XX
DE Banana bunchy top virus related nucleotide sequence #2.
XX
KW Banana bunchy top virus; BBTV; detection; ds.
XX
OS Banana bunchy top virus.
XX
PN TW360710-A.
XX
PD 11-JUN-1999.
XX
PF 30-JUN-1994; 94TW-0106105.
XX
PR 30-JUN-1994; 94TW-0106105.
XX
PA (BIOT-) DEV CENT BIOTECHNOLOGY.
XX
PI Wu R, You L, Song T;
XX
DR WPI; 2000-316145/27.
XX
XX Two circular single-stranded DNAs associated with banana bunchy top
PT virus and detection of the virus -
XX
XX Claim 2; Page 1; 7pp; Chinese.

XX The present invention describes two circular single-stranded DNAs
CC associated with banana bunchy top virus (BBTV) and the encoded proteins.
CC Also described is a method of detecting the BBTV virus using the
CC polymerase chain reaction (PCR). The present sequence represents a
CC BBTV related nucleotide sequence from the present invention.
XX
SQ Sequence 1091 BP; 345 A; 224 C; 247 G; 275 T; 0 other;
XX

Query Match 25.2%; Score 70.8; DB 21; Length 1091;
Best Local Similarity 56.2%; Pred. No. 4.1e-13;
Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;
XX

QY 8 ggtggaaccagaaagcgtattgtaccagtcacatcgaggaccacaaacgaaatttaac 67
DB 663 ggtggaagcgtcggtatgatgcacatcaaacgagtgatcctgataatcattggatt 722
QY 68 ctccggtaccaggtgtaatttagagttattaaattatgcctgttagaattgtttaag 127
DB 723 attgataccccagagtcattcagattatctgaattatggtggttatagacaaattaag 782
QY 128 acagggcattcagttcgacaaatacgaacc---ccttagttatcttgggttcaccat 184
DB 783 aatagagtttttaataataacaaatacgaaccatggtgattagaaaagatggacaaaat 842
QY 185 gtccatgtactcgtatttgcgaatgtcctgcctgattatttgaatacagcaggacaga 244
DB 843 gtccatgtattgttaggcaaatgtgtgctgattattgttaaaatttcagaagataga 902
QY 245 ataaactgtggaattatttaagtagtattgtgtcattc 278
DB 903 ataaaaataattattgttgagaaggaacttc 936

RESULT 3
AAV71833
ID AAV71833 standard; DNA; 1096 BP.
XX
AC AAV71833;
XX
DT 10-FEB-1999 (first entry)
XX
DE BBTV DNA II clone (2-17) nucleotide sequence.
XX
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
XX
OS Banana bunchy top virus.
XX

Key Location/Qualifiers
FH TATA_signal 1..7
FT /*tag= a
FT stem_loop 8..38
FT /*tag= b
FT CDS 70..928
FT /*tag= c
FT /*product= "ORF-V2 product"
FT /*transl_except= (pos:215..217, aa:Gly)
FT polyA_signal 533..538
FT /*tag= d
FT polyA_signal 799..804
FT /*tag= e
FT polyA_signal 907..912
FT /*tag= f
FT polyA_signal 1030..1035
FT /*tag= ge
XX
XX US5846705-A.
XX
XX 08-DEC-1998.
XX
XX 06-APR-1995; 95US-0418071.
XX


```

XX 06-APR-1995; 95US-0418071.
XX (BIOT-) DEV CENT BIOTECHNOLOGY.
XX Soong T, Wu R, You L;
XX WPI; 1999-059037/05.
XX P-PSDB; AAW87460.
XX
XX Nucleic acids having banana bunchy top virus component sequences -
XX used to design primers for use in polymerase chain reaction
XX detection of the virus
XX
XX Claim 2; Columns 17-18; 27pp; English.
XX
XX This represents the nucleotide sequence of a banana bunchy top virus
XX (BBTV) DNA II (clone 2-17) circular single stranded (css) DNA. The
XX invention provides nucleic acid sequences associated with BBTV that can
XX be used in a PCR technique for detecting BBTV. The nucleic acid
XX sequences (AAV71830 to AAV71833) are used as the basis for the
XX construction of PCR primers, to detect BBTV infection. The PCR technique
XX is used for detecting BBTV in plant tissues (preferably banana,
XX especially Musa species). The virus, one of the most important banana
XX species viruses, causes phloem damage and is transmitted by aphids. PCR
XX detection gives accurate, reliable and specific determination of absence
XX or presence of the virus.
XX
XX Sequence 1096 BP; 347 A; 231 C; 244 G; 274 T; 0 other;
XX
XX
XX Query Match 25.2%; Score 70.8; DB 20; Length 1096;
XX Best Local Similarity 56.2%; Pred. No. 4.1e-13;
XX Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;
XX
XX QY 8 ggtggaaccaggaagcgtattgtaccagtcacatgcagagcccaacgaaatttaac 67
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 668 ggtggaagcgtcgatgatgcacatcacaacgacgatgcctgataatcattgatt 727
XX QY 68 ctgcatgccaggtgtaattagagatttaaatatgccctgttagaattgttaag 127
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 728 attgatccccgaagtcattcagattatctgaattatggcgttatagacaataaag 787
XX QY 128 aacaggcattcagttcggacaataacgaacc---ccttagttatctgggttcgacct 184
XX Db || ||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 788 aatagagtttaataatacaataacgaacacatgctgtgattagaaagatggacaaat 847
XX QY 185 gtgcgtactcgtatttgcgaatgctcgtcgtgatttgaataatcagcaggacaga 244
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 848 gtccatgtaattgtatggcaatgtgtcgtgatttgaataatcagcaggacaga 907
XX QY 245 ataaaactgtggaatttaaaagtatgtgcac 278
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 908 ataaaataaattgttgagaaaggaaacttc 941
XX
XX RESULT 4
XX AAA38950
XX ID AAA38950 standard; DNA; 1096.BP.
XX
XX AC AAA38950;
XX
XX DT 23-AUG-2000 (first entry)
XX
XX DE Banana bunchy top virus related nucleotide sequence #5.
XX
XX KW Banana bunchy top virus; BBTV; detection; ds.
XX
XX OS Banana bunchy top virus.
XX
XX PN TW360710-A.
XX
XX PD 11-JUN-1999.

```

```

XX 30-JUN-1994; 94TW-0106105.
XX
XX 30-JUN-1994; 94TW-0106105.
XX
XX (BIOT-) DEV CENT BIOTECHNOLOGY.
XX
XX Wu R, You L, Song T;
XX
XX WPI; 2000-316145/27.
XX
XX Two circular single-stranded DNAs associated with banana bunchy top
XX virus and detection of the virus -
XX
XX Claim 2; Page 1; 7pp; Chinese.
XX
XX The present invention describes two circular single-stranded DNAs
XX associated with banana bunchy top virus (BBTV) and the encoded proteins.
XX Also described is a method of detecting the BBTV virus using the
XX polymerase chain reaction (PCR). The present sequence represents a
XX BBTV related nucleotide sequence from the present invention.
XX
XX Sequence 1096 BP; 347 A; 231 C; 244 G; 274 T; 0 other;
XX
XX
XX Query Match 25.2%; Score 70.8; DB 21; Length 1096;
XX Best Local Similarity 56.2%; Pred. No. 4.1e-13;
XX Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;
XX
XX QY 8 ggtggaaccaggaagcgtattgtaccagtcacatgcagagcccaacgaaatttaac 67
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 668 ggtggaagcgtcgatgatgcacatcacaacgacgatgcctgataatcattgatt 727
XX QY 68 ctgcatgccaggtgtaattagagatttaaatatgccctgttagaattgttaag 127
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 728 attgatccccgaagtcattcagattatctgaattatggcgttatagacaataaag 787
XX QY 128 aacaggcattcagttcggacaataacgaacc---ccttagttatctgggttcgacct 184
XX Db || ||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 788 aatagagtttaataatacaataacgaacacatgctgtgattagaaagatggacaaat 847
XX QY 185 gtgcgtactcgtatttgcgaatgctcgtcgtgatttgaataatcagcaggacaga 244
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 848 gtccatgtaattgtatggcaatgtgtcgtgatttgaataatcagcaggacaga 907
XX QY 245 ataaaactgtggaatttaaaagtatgtgcac 278
XX Db ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
XX 908 ataaaataaattgttgagaaaggaaacttc 941
XX
XX RESULT 5
XX AAV71832
XX ID AAV71832 standard; DNA; 1106 BP.
XX
XX AC AAV71832;
XX
XX DT 10-FEB-1999 (first entry)
XX
XX DE BBTV DNA I clone (7-4-2) nucleotide sequence.
XX
XX KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
XX Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
XX
XX OS Banana bunchy top virus.
XX
XX Key Location/Qualifiers
XX TATA_signal 1..8
XX stem_loop /*tag= a
XX /*tag= b
XX CDS 62..922
XX /*tag= c
XX /product= "ORF-V2 product"

```



```

DT 23-AUG-2000 (first entry)
XX
DE Banana bunchy top virus related nucleotide sequence #1.
XX
KW Banana bunchy top virus; BBTV; detection; ds.
XX
OS Banana bunchy top virus.
XX
PN TW360710-A.
XX
PD 11-JUN-1999.
XX
PF 30-JUN-1994; 94TW-0106105.
XX
PR 30-JUN-1994; 94TW-0106105.
XX
PA (BIOT-) DEV CENT BIOTECHNOLOGY.
XX
PI Wu R, You L, Song T;
XX
XX WPI; 2000-316145/27.
XX
DR Two circular single-stranded DNAs associated with banana bunchy top
PT virus and detection of the virus
XX
PS Claim 1; Page 1; 7pp; Chinese.
XX
CC The present invention describes two circular single-stranded DNAs
CC associated with banana bunchy top virus (BBTV) and the encoded proteins.
CC Also described is a method of detecting the BBTV virus using the
CC polymerase chain reaction (PCR). The present sequence represents a
CC BBTV related nucleotide sequence from the present invention.
XX
SQ Sequence 1106 BP; 335 A; 226 C; 256 G; 289 T; 0 other;

Query Match 24.6%; Score 69.2; DB 21; Length 1106;
Best Local Similarity 54.8%; Pred. No. 1.3e-12;
Matches 137; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 20 aaggacgtattgtaccagtagcatcgaggaccacacgaatttattcctcgtatgcc 79
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 677 aatatcttctcctatgtggcagagtagtcgacagcatatagtattgtattcct 736
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 80 aggtgtaatttagagattttaaattatgcctgttagaattgtttgaacacaggcattc 139
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 737 cctgttaacaggattattttaaattatgatgtaataagaggcattaaagggttata 796
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 140 agttcgacaaatacgaaccccttagttatcttgggttcgacacatgtgcatgtactcgt 199
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 797 gagagtactaaatacaaacacccataagatagttgaattaggttaaaatacatgtatcgtc 856
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 200 ttgccaatgtcctgcgtgatttttgaataacagcagggacagataaaactgtggaat 259
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 857 atggcgaatttcactgactctgttaaaatcgcgaagatcgaataaaatcatttat 916
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 260 atttaaga 269
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 917 tgctgaaga 926
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AAX34686
ID AAX34686 standard; DNA; 593 BP.
XX
AC AAX34686;
XX
DT 02-JUL-1999 (first entry)
XX
DE DNA sequence of BBTV S1 promoter fragment including an ORF fragment.
XX
KW Promoter; Banana Bunchy Top Virus; BBTV; gene transcription; ubil;
KW polyubiquitin 1; banana; ss.

XX OS Banana bunchy top virus.
XX PN WO915646-A1.
XX PD 01-APR-1999.
XX PF 21-SEP-1998; 98WO-AU00786.
XX PR 30-JUN-1998; 98AU-0004423.
XX PR 19-SEP-1997; 97AU-0009339.
XX PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
XX PI Becker DK, Dale JL, Dugdale B, Harding RM, Hermann SR;
XX WPI; 1999-254706/21.
XX
XX Banana Bunchy Top Virus promoter constructs
XX
XX Examples; Fig 22; 84pp; English.
XX
CC The invention relates to DNA promoter sequences derived from components
CC of Banana Bunchy Top Virus (BBTV). The DNA molecule includes a promoter
CC sequence derived from an untranslated portion of any one BBTV components
CC and is adaptable for promoting transcription of a cloned gene in a plant
CC cell. The promoter, included in a DNA chimeric vector, is useful for the
CC expression of a gene in a plant cell. The inclusion of the polyubiquitin
CC 1 (ubil) intron sequence into BBTV promoter constructs enhances promoter
CC activity. The present sequence represents the DNA sequence of S1 promoter
CC fragment derived from BBTV-S1 and includes an open reading frame fragment
CC and an intergenic region fragment.
XX
SQ Sequence 593 BP; 189 A; 103 C; 133 G; 168 T; 0 other;

Query Match 24.0%; Score 67.4; DB 20; Length 593;
Best Local Similarity 55.8%; Pred. No. 3.9e-12;
Matches 150; Conservative 0; Mismatches 116; Indels 3; Gaps 1;

QY 1 cacatgtgtggaaacagagacgtattgtaccagtagcat---cgaggaccacaaaacg 57
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 73 caccagagtgaggaaagagagacatactgtctctacgtggacgagagatctgagaa 132
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 58 aaatttaacctcgtatgccaggtgtaatttagattattaaattatgcctgttaga 117
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 133 gcattattgtattgtattcctcgtgttaatacagattatttaaattatgtattaga 192
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 118 atgtgttaagacagggcattcagtcggaacaaatacgaaccccttacttcttggtt 177
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 ggcattaaaggatagggtgatagagagactaataataaccccttacttcttggtt 252
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 cgaccatgtcatgtactcgtattgttgcattgtcctgctgattatttgaataacagcag 237
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 253 gattaataacatgaattgtcattgtcgttaatttcagtcagaaattctgtaaaatccga 312
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 gacagaataaaactgtggaattttaa 266
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 313 agatagaataaagattattttgttaa 341
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AAT13165
ID AAT13165 standard; DNA; 1017 BP.
XX
AC AAT13165;
XX
DT 23-MAY-1996 (first entry)
XX
DE SCSV segment 6.
XX
KW SCSV; promoter; transcription; transgenic plant; legume;
KW gene expression; crop improvement; ss.

```

```

XX OS Subterranean clover stunt virus isolate F.
XX FH
XX FT Location/Qualifiers
XX FT 48..905
XX FT /*tag= a
XX PN WO9606932-Al.
XX PD
XX PD 07-MAR-1996.
XX PF 30-AUG-1995; 95WO-AU00552.
XX PR 07-NOV-1994; 94AU-0009281.
XX PR 30-AUG-1994; 94AU-0007770.
XX PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX PA (AUSU ) UNIV AUSTRALIAN NAT.
XX PI Boevink PC, Chu PWG, Keese PK, Khan RL, Larkin PJ;
XX PI Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
XX DR WPI; 1996-160363/16.
XX PT Circovirus transcription regulatory sequences and related constructs
XX PT - useful in plants, esp. leguminous plants, for the modulation of
XX PT gene expression
XX PS Claim 4; Page 62; 121pp; English.
XX CC The subterranean clover stunt virus (SCSV) genome has at least 7
XX CC distinct ssDNA components, designated segments 1-7 (AAT13160-66), each
XX CC contg. 1 major open reading frame and a non-coding region. Segment
XX CC 6 is predicted to be a viral replication-associated protein gene.
XX CC Genetic constructs useful in the genetic engineering of plants (esp.
XX CC legumes) comprise 1 or more heterologous gene(s) operatively linked
XX CC to a promoter region, and in some cases also a terminator region,
XX CC selected from segments 1-7. The transcription regulators facilitate
XX CC expression of foreign genes in plants and also facilitate control of
XX CC levels of gene expression in different plant tissue types.
XX SQ Sequence 1017 BP; 312 A; 160 C; 265 G; 280 T; 0 other;

Query Match 23.2%; Score 65.2; DB 17; Length 1017;
Best Local Similarity 55.2%; Pred. No. 2,5e-11;
Matches 149; Conservative 0; Mismatches 118; Indels 3; Gaps 1;

QY 7 tggatgaacacagaaaggacgtattgtaccagtcacatcgaggaccacaaacacgaatttaatt 66
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 644 tggagaaagacacagatgatgtcatctgtgactgtgagcctgaagaataattgggt 703

QY 67 cctcgatgtaccagtgtaattagagatttaattatgcctgtttagaattgtttaa 126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 704 attgacataccacagattagttcagagtgatgtaattatgtgtgaatagaacaggttaa 763

QY 127 gaacagggtatcagttcggaacaaatcacacaccccttagttatcttgggtcgacca--- 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 764 gaatagggttaattggtgaactaactagatgagccatgtgtaatgcggatgataatcatcc 823

QY 184 tgtgcattactcgtatttgccaatgtcctgcctgattatttgaaatcagcaggagacag 243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 824 ttttcattgaattgtgttgcaaatgtactccagatttgggaaattaaatgaagatag 883

QY 244 aataaactgtggaattatttaagtgtg 273
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 884 aataaataattcgttgttgaaactctg 913

RESULT 10
AAT13161
ID AAT13161 standard; DNA; 1022 BP.
XX

```

```

AC AAT13161;
XX
XX DT 23-MAY-1996 (first entry)
XX
XX DE SCSV segment 2.
XX
XX FT SCSV; promoter; transcription; transgenic plant; legume;
XX FT gene expression; crop improvement; ss.
XX
XX OS Subterranean clover stunt virus isolate F.
XX
XX FH Key Location/Qualifiers
XX FH 82..924
XX FT CDS /*tag= a
XX
XX PN WO9606932-Al.
XX
XX PD 07-MAR-1996.
XX
XX PF 30-AUG-1995; 95WO-AU00552.
XX
XX PR 07-NOV-1994; 94AU-0009281.
XX PR 30-AUG-1994; 94AU-0007770.
XX
XX PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX PA (AUSU ) UNIV AUSTRALIAN NAT.
XX PI Boevink PC, Chu PWG, Keese PK, Khan RL, Larkin PJ;
XX PI Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
XX
XX WPI; 1996-160363/16.
XX
XX DR Circovirus transcription regulatory sequences and related constructs
XX DR - useful in plants, esp. leguminous plants, for the modulation of
XX DR gene expression
XX
XX PS Claim 4; Page 58; 121pp; English.
XX
XX CC The subterranean clover stunt virus (SCSV) genome has at least 7
XX CC distinct ssDNA components, designated segments 1-7 (AAT13160-66), each
XX CC contg. 1 major open reading frame and a non-coding region. Segment
XX CC 2 is predicted to be a viral replication-associated protein gene.
XX CC Genetic constructs useful in the genetic engineering of plants (esp.
XX CC legumes) comprise 1 or more heterologous gene(s) operatively linked
XX CC to a promoter region, and in some cases also a terminator region,
XX CC selected from segments 1-7. The transcription regulators facilitate
XX CC expression of foreign genes in plants and also facilitate control of
XX CC levels of gene expression in different plant tissue types.
XX SQ Sequence 1022 BP; 309 A; 183 C; 245 G; 285 T; 0 other;

Query Match 22.0%; Score 61.8; DB 17; Length 1022;
Best Local Similarity 52.1%; Pred. No. 3.1e-10;
Matches 138; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 2 acatgtgttggaacacagaaaggacgtattgtaccagtcacatcgaggaccacaaacgaatt 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 661 acagcggagggaagaccacagcgtattatgtatgtatgtatgtatgtatgtatgtatgtat 720

QY 62 ttaatcctcgtatgtaaccaggtgtaatttagagatttaattatccctgttagaattgt 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 721 attcgtgttgatgtctccaggtgtctctcggagatgtagaactatcaggcgatggagatg 780

QY 122 gttagaacacaggcattcagttcggacaataacacaccccttagttatcttgggttcgac 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 781 ttgaagaacagagtttttgcgaataataatagcctgtagatctttagtaggaag 840

QY 182 catgtgcatgtactcgtatttggccaatgtcctgcctgattatttgaaatcagcaggagac 241
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 841 ttagtcttaattgtgttgcacacgtggcactgacccacgcgcataagtggagac 900

QY 242 agaataaaactgtggaattatttaa 266

```

Db 901 agacttgtaattatcaattgtttaa 925

RESULT 11

AAV24093
ID AAV24093 standard; DNA; 1105 BP.

ID AAV24093 standard; DNA; 1105 bp.

XX

AC AAV24093;

XX

DT 11-AUG-1998 (first entry)

[illegible]

Banana bunchy top virus

XX
XX
XX

KW BBTV: probe: diagnostic probe

XX
XX
xxiv, probe; diagnosis etc print

Banana bunchy top virus

BANANA BUNCHY TOP VIRUS.

XX

XX

1155756708-A
XX
PN

FN
XX
US3/36/08-A.

XX
PD
26-MAY-1998

FD
20 MAY 1998.
XX

XX
DE
34-FEB-1994 04:03:00

24-FEB-1994; 94US-0202186.

XX
PP
24-FEB-1964
0430.0000000

PR 24-FEB-1994; 94US-0202186.

Category	Item	Value
A	1	100
	2	100
B	1	100
	2	100
C	1	100
	2	100
D	1	100
	2	100
E	1	100
	2	100
F	1	100
	2	100
G	1	100
	2	100
H	1	100
	2	100
I	1	100
	2	100
J	1	100
	2	100
K	1	100
	2	100
L	1	100
	2	100
M	1	100
	2	100
N	1	100
	2	100
O	1	100
	2	100
P	1	100
	2	100
Q	1	100
	2	100
R	1	100
	2	100
S	1	100
	2	100
T	1	100
	2	100
U	1	100
	2	100
V	1	100
	2	100
W	1	100
	2	100
X	1	100
	2	100
Y	1	100
	2	100
Z	1	100
	2	100

(UYQU-) UNIV QUEENSLAND TECH

XX

PI Burns TM, Dale JL, Harding RM, Karan M;

XX

WPI; 1998-321636/28.

Isolated Banana bunchy top virus DNA - useful as diagnostic probes

and primers and for producing virus-resistant plants

XX
XX

PS Disclosure; Column 29-30; 59pp; English.

XX

CC This sequence represents

the B

dia

organisms e.g. to produce virus.

enhancers

XX
XXXXXXXXXXXX

```

XX PD 01-FEB-2001.
XX PF
XX PR 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT a single surface
XX PS Example 6; Page 127; 159pp; English.
XX SS The present invention relates to a composition comprising two nucleic
XX CC acids each containing an electron-transfer group (ETM) having
XX CC different redox potentials. The invention is used for electronic
XX CC detection of nucleic acids, especially of substitutions (mismatches)
XX CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX CC monitoring gene expression.
XX SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 15.6%; Score 43.8; DB 22; Length 936;
Best Local Similarity 2.4%; Pred. No. 0.00018;
Matches 6; Conservative 156; Mismatches 93; Indels 0; Gaps 0;

QY 27 tatgtaccagtcacatgcagagcccaaaacgaatttaattcctcgatgtaccaggtgta 86
Db 600 WWWWWWCCCCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 541
QY 87 attagagatttaattatgcctgttagaattgtttaagaacagggcattcagttcgg 146
Db 540 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 481
QY 147 acaaatcacgaaccccttagtcttgcgtgacccatgtgcatgtactcgtattgcca 206
Db 480 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 421
QY 207 atgtcctgcctgattatttgaaatcagcagggacagaataaaactgtggaattataa 266
Db 420 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWW 361
QY 267 gtatgtgtcatctaa 281
Db 360 WWWWWWWWWWWWWWW 346

RESULT 14
AAF58254/C
ID AAF58254 standard; DNA; 936 BP.
XX AC AAF58254;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1875.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN W0200107665-A2.
XX PD 01-FEB-2001.
XX PF

```

```

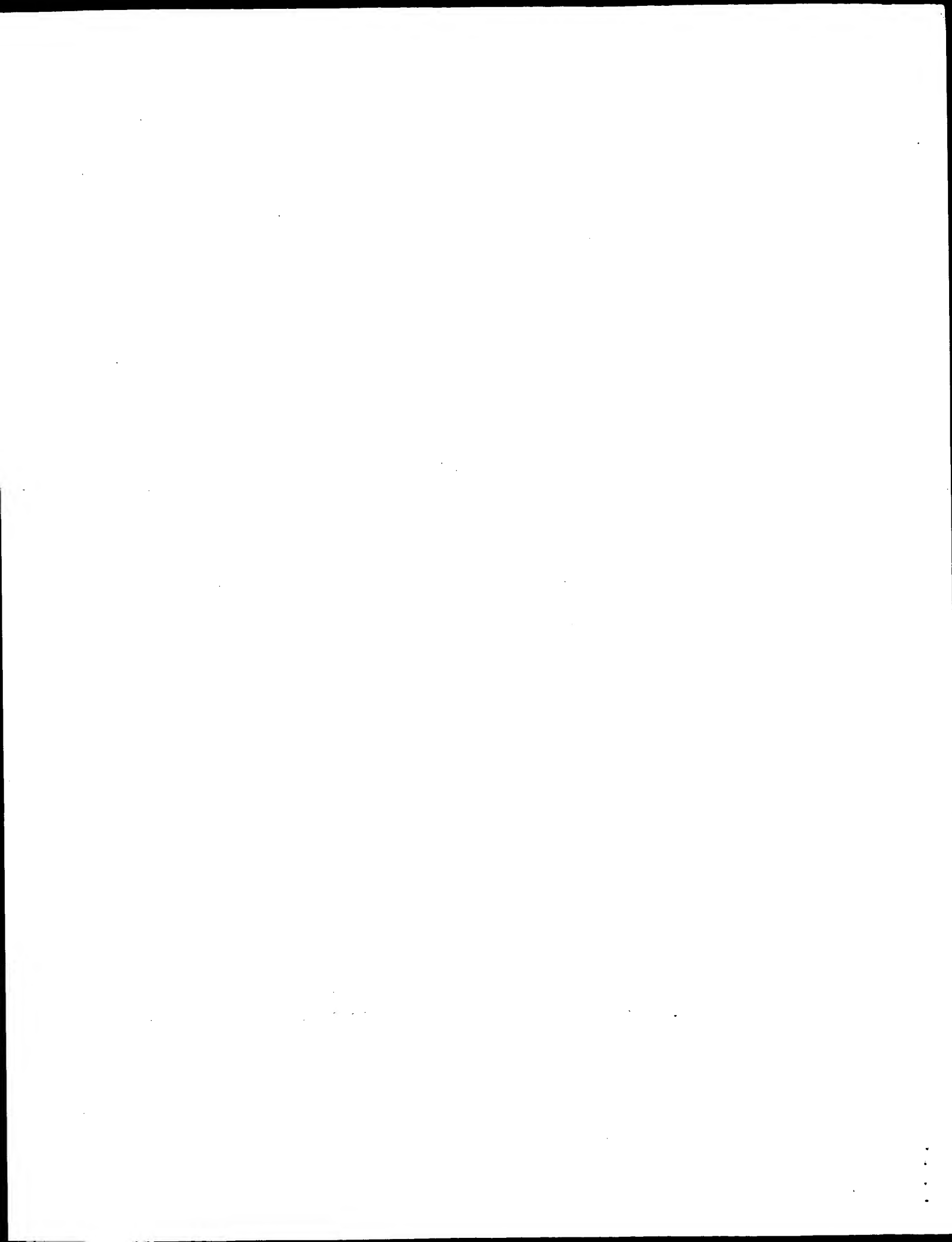
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT a single surface
XX PS Example 6; Page 127; 159pp; English.
XX SS The present invention relates to a composition comprising two nucleic
XX CC acids each containing an electron-transfer group (ETM) having
XX CC different redox potentials. The invention is used for electronic
XX CC detection of nucleic acids, especially of substitutions (mismatches)
XX CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX CC monitoring gene expression.
XX SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 15.6%; Score 43.8; DB 22; Length 936;
Best Local Similarity 2.4%; Pred. No. 0.00018;
Matches 6; Conservative 156; Mismatches 93; Indels 0; Gaps 0;

QY 27 tatgtaccagtcacatgcagagcccaaaacgaatttaattcctcgatgtaccaggtgta 86
Db 600 WWWWWWCCCCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 541
QY 87 attagagatttaattatgcctgttagaattgtttaagaacagggcattcagttcgg 146
Db 540 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 481
QY 147 acaaatcacgaaccccttagtcttgcgtgacccatgtgcatgtactcgtattgcca 206
Db 480 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 421
QY 207 atgtcctgcctgattatttgaaatcagcagggacagaataaaactgtggaattataa 266
Db 420 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWW 361
QY 267 gtatgtgtcatctaa 281
Db 360 WWWWWWWWWWWWWWW 346

RESULT 15
AAF58257/C
ID AAF58257 standard; DNA; 936 BP.
XX AC AAF58257;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1954.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN W0200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search., using sw model

Run on: October 24, 2001, 13:13:47 ; Search time 5479.82 Seconds
(without alignments)
484.733 Million cell updates/sec

Title: us-09-462-955-1_copy_711_991

Perfect score: 281

Sequence: 1 cacatgtgtggaaccagaa.....ttaaagtatgtgtatctaa 281

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estin6.*
69: em_estin7.*
70: em_estin8.*
71: em_estin9.*
72: em_estin10.*
73: em_estin11.*
74: em_estin12.*
75: em_estin13.*
76: em_estin14.*
77: em_estin15.*
78: em_estin16.*
79: em_estin17.*
80: em_estin18.*
81: em_estin19.*
82: em_estin20.*
83: em_estin21.*
84: em_estin22.*
85: em_estin23.*
86: em_estin24.*
87: em_estin25.*
88: em_estin26.*
89: em_estin27.*
90: em_estin28.*
91: em_estin29.*
92: em_estin30.*
93: em_estin31.*
94: em_estin32.*
95: em_estin33.*
96: em_estin34.*
97: em_estin35.*
98: em_estin36.*
99: em_estin37.*
100: em_estin38.*
101: em_estin39.*
102: em_estin40.*
103: em_estin41.*
104: em_estin42.*
105: em_estin43.*
106: em_estin44.*
107: em_estin45.*
108: em_estin46.*
109: em_estin47.*
110: em_estin48.*
111: em_estin49.*
112: em_estin50.*
113: em_estin51.*
114: em_estin52.*
115: em_estin53.*
116: em_estin54.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Query Match	12.7%;	Score	35.6;	DB	224;	Length	318;
Best Local Similarity	54.6%;	Pred.	No. 3.6;				

```
Seq primer: puc 18 forward:
High quality sequence start: 76
High quality sequence stop: 183.
Location/Qualifiers
    1..518
Source
    genome="Homo sapiens"
```


12.5%; Score 35.2; DB 233; Length 848;
ty 55.8%; Pred. No. 5.6;
ervative 0; Mismatches 53; Indels 0; Gaps

GTCCCTGTTGCAGATGACATGATGTTGTATATTTAGAAATCCCCCATGGC 455

487 bp DNA GSS 23-MAR-1999
0318.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203

1 61:4476249

ens ; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom
Eutheria; Primates; Catarrhini; Hominidae; Homo.

Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Ventel

IC End Sequences from Library RPCI-11 for Sequence-Read

ing
ed (1997)
s: RPCI-11-20318 TJ

Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
Institute for Genomic Research

Institute for Genomic Research
 Medical Center Dr., Rockville, MD 20850
 838 0200

838 0208
peetiqr.org

are derived from the human BAC library RPCI-11. For BAC availability, please contact Pieter de Jong

dejong.med.buffalo.edu). Clones may be purchased from resources (<http://bacpac.med.buffalo.edu/ordering>) or from (infoaragon.com). PAC and search page:

Genet cs (inforesgen.com): bac_end_search.page.
www.tigr.org/tdb/humgen/bac_end_search.ht
er. T7

Location/Qualifiers

```
1. .487
/organism="Homo sapiens"
```

```

/db_xref="GDB:7577767"
/db_xref="taxon:9606"

```

```
/clone="RPCI-11-20318"  
/clone_lib="RPCI-11"  
/cov="Malign"
```

```

/sex= male
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI"

```

RPC111 Human Male BAC Libraryⁿ

12.3%; Score 34.6; DB 228; Length 487;

ity 54.3%; Pred. NO. 7.0;
servative 0; Mismatches 59; Indels 0; Gaps

atgtgttaagaacagggcatttcagttcggacaatacgaaccccttagtt 167

```

Db 127 CTCTGTAGTAATTCGGTCCTCCAGGCAATCAGCAGGAGAAATAAAGGGTGTTC 186
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 168 attctgggttcaccatgtgcatgtactgtattgttgcccaatgctcgtctgattttga 227
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 AATTAGAAAAGGAAGTCAAAATTTGTCCTGTTTGCAGATCACATGATTGTCATCTAC 246
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 228 aatcagca 236
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 AAAACCCCA 255
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
LOCUS A0417598 555 bp DNA GSS 23-MAR-1999
DEFINITION RPCI-11-203C8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-203C8,
DNA sequence.
ACCESSION A0417598
VERSION A0417598.1 GI:4475441
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 555)
AUTHORS Zhao,S., Adams,W.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: RPCI-11-203C8.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..555
                     /organism="Homo sapiens"
                     /db_xref="GDB:7577623"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-203C8"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT          192 a 114 c 110 g 139 t
ORIGIN
Query Match          12.3%; Score 34.6; DB 228; Length 555;
Best Local Similarity 54.3%; Pred. No. 7.8;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 108 ccctgttagaattgtttaagcaggcatttcagtttcgacaaatacgaaccccttagtt 167
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 CTCTGGTAGAATTCGGTCCTCCAGGCAATCAGCAGGAGAAATAAAGGGTGTTC 209
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 168 attctgggttcaccatgtgcatgtactgtattttgcccaatgctcgtctgattttga 227
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 210 AATTAGAAAAGGAAGTCAAAATTTGTCCTGTTTGCAGATCACATGATTGTCATCTAC 269
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 228 aatcagca 236
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270 AAAACCCCA 278
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
LOCUS CNS016H0 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACNI6B20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106734.1 GI:5623558
VERSION AL106734.1
KEYWORDS fruit fly.
SOURCE Plasmid Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelobAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACNI6B20"
                     /note="end : T7"
BASE COUNT          260 a 123 c 159 g 217 t 342 others
ORIGIN
Query Match          12.3%; Score 34.6; DB 219; Length 1101;
Best Local Similarity 30.8%; Pred. No. 8.9;
Matches 57; Conservative 47; Mismatches 81; Indels 0; Gaps 0;

QY 90 taqagtatttaattatgccctgttagaattgtttaagaaacaggcatttcaggaca 149
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 666 TAAATWADTWATWATWBTAAATTTTNNASAAARTTWGGGAKADWTTTATTKRGR 725
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 150 aatacgaaccccttagttatcttggttcgaccatgtgcactgtactgtatttgcgaatg 209
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 726 DGGGGCAAGACAGYKATADTATATGKAKRSGAARGGWKGTGTRTATATWGGDRKAGW 785
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 tctgcctgattatttgaaataacacagcaggagcagataaaactgtggaattttaaagta 269
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 786 TTGTTKTKTKTKAWCKTDKAKTGAARTGAGGGGAWBGARAVTGWGATWTDABWTSTDWRCAG 845
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 270 ttgtt 274
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 846 NRTAT 850
      ||| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
LOCUS A0010299 443 bp DNA GSS 21-MAY-1998
DEFINITION HS-2172_B1_MR_G08 ClT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2172 Col=15 Row=N, DNA sequence.
ACCESSION A0010299
VERSION A0010299.1 GI:3143369

```



```

BASE COUNT      218 a   98 c   138 g   150 t
ORIGIN

Query Match      12.0%; Score 33.8; DB 219; Length 604;
Best Local Similarity 62.4%; Pred. No. 14;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 50 ccaaacgaaatttaactcgtatgaccaggtgttaattagagatttttaataatgcc 109
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 CTAAACAAATGAAGTCATCTTAACACAGGTGAATTTATAGTGTGCAATTAATA 370
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 110 ctgttagaattgtttaagaacagg 134
    || || || || || || || || || || || || || || || || || || ||
Db 371 CTAAATAAAACTGTTTAAAAAAAAGG 395

RESULT 14
A0488982      772 bp   DNA      GSS      24-APR-1999
LOCUS      RPCI-11-267P13.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-267P13
DEFINITION      , DNA sequence.
ACCESSION      A0488982
VERSION      A0488982.1 GI:4674856
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 772)
JOURNAL      Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter
COMMENT      J.C.
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
Unpublished (1997)
Other_GSSs: RPCI-11-267P13.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@igr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pleter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.
FEATURES
    source
        Location/Qualifiers
            1..772
                /organism="Homo sapiens"
                /db_xref="GB:7602516"
                /db_xref="taxon:9606"
                /clone="RPCI-11-267P13"
                /clone_lib="RPCI-11"
                /sex="Male"
                /cell_type="Lymphocytes"
                /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPCI11 Human Male BAC Library"
BASE COUNT      300 a   172 c   122 g   178 t
ORIGIN

Query Match      12.0%; Score 33.8; DB 229; Length 772;
Best Local Similarity 53.4%; Pred. No. 14;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 122 gtttaagaacggcattcagtcggacaaatcgaaccccttagttatcttggttcgac 181
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 GTTCTGTCAGGCATCAGGCAGGAAGAATAACGGATATTCAGTTAGGAAAAGAG 573

BASE COUNT      218 a   98 c   138 g   150 t
ORIGIN

Query Match      12.0%; Score 33.8; DB 221; Length 894;
Best Local Similarity 60.9%; Pred. No. 15;
Matches 53; Conservative 1; Mismatches 33; Indels 0; Gaps 0;

QY 69 tcatgtaccaggtgtaattagagattttaaaatagccctgttaaatgtgttaaga 128
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 793 TCGAAGCACACGTTTACTTTCTTAATAAATATACATGCTTGGAGTAGGTGTGCAGA 852
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 129 acaggcattcagttccggacaaatcac 155
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 853 ACAGAGCGTKATGTCGGATGTATAAG 879
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: October 24, 2001, 13:13:48

```

Job time: 11813 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:27 ; Search time 180.6 Seconds
(without alignments)
294.554 Million cell updates/sec

Title: US-09-462-955-1_COPY_711_991
Perfect score: 281
Sequence: 1 cacatgtgtggaacagaa.....ttaaagtatgtcatctaa 281

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.8	25.2	1091	2	US-08-418-071-5
2	70.8	25.2	1096	2	US-08-418-071-4
3	70.8	25.2	1106	2	US-08-418-071-3
4	65.2	23.2	1017	4	US-08-793-634B-6
5	61.8	22.0	1022	4	US-08-793-634B-2
6	44.6	15.9	1105	1	US-08-202-186-18
7	44.4	15.8	1103	1	US-08-202-186-16
8	43	15.3	1110	1	US-08-202-186-14
9	43	15.3	1111	1	US-08-202-186-12
10	42.8	15.2	1104	1	US-08-202-186-17
11	41.4	14.7	982	3	US-08-973-068-28
12	41.4	14.7	1109	1	US-08-202-186-13
13	41.4	14.7	1111	1	US-08-202-186-9
14	41.4	14.7	1111	1	US-08-202-186-24
15	40	14.2	1111	1	US-08-202-186-10
16	39.8	14.2	1110	1	US-08-202-186-11
17	39.8	14.2	1111	1	US-08-202-186-15
18	29.4	10.5	8920	2	US-08-446-855A-1
19	29.4	10.5	8920	4	US-09-150-741-1
20	28.8	10.2	10607	1	US-08-078-090-3
21	27.8	9.9	2570	2	US-09-056-075-2
22	27.6	9.8	2899	2	US-08-624-581-2
23	27.6	9.8	2917	2	US-08-624-581-3
24	27.6	9.8	2959	2	US-08-624-581-1
25	27.6	9.8	6476	4	US-09-127-670-5
26	27.4	9.8	840	5	PCT-US91-08177-12
27	27.4	9.8	2256	6	5220013-1

c 28	27.4	9.8	7323	5	PCT-US91-08177-1	Sequence 1, Appli
c 29	27.2	9.7	5892	3	US-08-755-587-27	Sequence 27, Appl
c 30	27.2	9.7	7240	3	US-08-755-587-15	Sequence 15, Appl
c 31	27.2	9.7	11283	2	US-08-603-753D-3	Sequence 3, Appli
c 32	27.2	9.7	11283	3	US-09-099-753-3	Sequence 3, Appli
c 33	27.2	9.7	11283	4	US-08-986-106-3	Sequence 3, Appli
c 34	27.2	9.7	11283	4	US-08-986-106-3	Sequence 3, Appli
c 35	27.2	9.7	11385	2	US-08-639-501-1	Sequence 1, Appli
c 36	27.2	9.7	11385	3	US-09-044-946-1	Sequence 1, Appli
c 37	27	9.6	1815	4	US-09-586-935-2	Sequence 2, Appli
c 38	27	9.6	2042	4	US-08-933-821-16	Sequence 16, Appl
c 39	27	9.6	2042	3	US-08-934-494-5	Sequence 5, Appli
c 40	27	9.6	2042	3	US-08-960-507-16	Sequence 16, Appl
c 41	27	9.6	2042	3	US-08-143-068-5	Sequence 5, Appli
c 42	26.8	9.5	1320	2	US-08-853-659A-12	Sequence 12, Appl
c 43	26.8	9.5	15512	2	US-08-853-659A-5	Sequence 5, Appli
c 44	26.8	9.5	15512	2	US-08-853-659A-8	Sequence 8, Appli
c 45	26.8	9.5	15512	2	US-08-853-659A-63	Sequence 63, Appl

ALIGNMENTS

RESULT 1
US-08-418-071-5
; Sequence 5, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1091 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: subgenomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-418-071-5

Query Match 25.2%; Score 70.8; DB 2; Length 1091;
 Best Local Similarity 56.2%; Pred. No. 2.5e-14;
 Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

QY 8 ggtggaaccagaaagcgtattgtaccagttacatcgagagaccacaaacgaaatttaac 67
 DB 663 GGTGGAACAGCTGGATATGATGCACATCATACGATGGATCCTGATATCATTTGATT 722
 QY 68 ctgatgtaccaggtgtaatttagagattttaattatccctgttagaattgtttaag 127
 DB 723 ATTGATATCCCGAAGCAATTCAGATTATCTGAATTATGCGTTATAGAACAAATTAAG 782
 QY 128 aacagggttcagtcggaacaaatagaaac---ccttagttatcttgggttcgacct 184
 DB 783 AATAGAGTTTAAATAAATACAAAATACGAACCATGTGTGATGAAAAGATGGACAAAT 842
 QY 185 gtgcattgtactcgtatttgcgaatgtccctgctgatttgaataatcagcaggacaga 244
 DB 843 GTCCATGTAATTTGTTATGGCAAAATGTTGCTGATTTGTAATAATTTTCAAGAGATAGA 902
 QY 245 ataaactgtggaatatttaagattgtgtcatc 278
 DB 903 ATAAAAATAAATAATTTGTTGAGAAAGGAACTTC 936

RESULT 2
 US-08-418-071-4
 ; Sequence 4, Application US/08418071
 ; Patent No. 5846705
 ; GENERAL INFORMATION:
 ; APPLICANT: Wu, Rey-Yuh
 ; APPLICANT: You, Li-Ru
 ; APPLICANT: Soong, Tai-Seng
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECT
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10020
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 5.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/418,071
 ; FILING DATE: 06-APR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haley Jr., James F.
 ; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: DCB-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 596-9000
 ; TELEFAX: (212) 596-9090
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1096 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: genomic DNA
 ; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 ; US-08-418-071-4

Query Match 25.2%; Score 70.8; DB 2; Length 1096;
 Best Local Similarity 56.2%; Pred. No. 2.5e-14;
 Matches 154; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

QY 8 ggtggaaccagaaagcgtattgtaccagttacatcgagagaccacaaacgaaatttaac 67
 DB 668 GGTGGAACAGCTGGATATGATGCACATCATACGATGGATCCTGATATCATTTGATT 727
 QY 68 ctgatgtaccaggtgtaatttagagattttaattatccctgttagaattgtttaag 127
 DB 728 ATTGATATCCCGAAGCAATTCAGATTATCTGAATTATGCGTTATAGAACAAATTAAG 787
 QY 128 aacagggttcagtcggaacaaatagaaac---ccttagttatcttgggttcgacct 184
 DB 788 AATAGAGTTTAAATAAATACAAAATACGAACCATGTGTGATGAAAAGATGGACAAAT 847
 QY 185 gtgcattgtactcgtatttgcgaatgtccctgctgatttgaataatcagcaggacaga 244
 DB 848 GTCCATGTAATTTGTTATGGCAAAATGTTGCTGATTTGTAATAATTTTCAAGAGATAGA 907
 QY 245 ataaactgtggaatatttaagattgtgtcatc 278
 DB 908 ATAAAAATAAATAATTTGTTGAGAAAGGAACTTC 941

RESULT 3
 US-08-418-071-3
 ; Sequence 3, Application US/08418071
 ; Patent No. 5846705
 ; GENERAL INFORMATION:
 ; APPLICANT: Wu, Rey-Yuh
 ; APPLICANT: You, Li-Ru
 ; APPLICANT: Soong, Tai-Seng
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
 ; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
 ; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States
 ; ZIP: 10020
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 5.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/418,071
 ; FILING DATE: 06-APR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haley Jr., James F.
 ; REGISTRATION NUMBER: 27,794
 ; REFERENCE/DOCKET NUMBER: DCB-1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 596-9000
 ; TELEFAX: (212) 596-9090
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1106 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: genomic DNA
 ; DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-793-634B-2

Query Match      22.0%; Score 61.8; DB 4; Length 1022;
Best Local Similarity 52.1%; Pred. No. 2.4e-11;
Matches 138; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 2 acatgtgtggaacaggaagcgtatgtaccagtlacatcgaggaccacaaacgaat 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 661 ACAGCCGAGGAGAACCCAGGAGCTATTATATATATATATATATATATATATAT 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 ttaatcctcgatgtaccagtgtaatttagagtgatttaattatgccttgtagaagt 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 721 ATTGGCTTTGATGTTCCAGGCTGTTCTTCGGAGAGATGATGAACCTATCAGGCGATGGAGATG 780
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 122 gttaaagacaggcattcagttcggacaaatacgaaccccttagttctgggtcgac 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 781 TTGAAGAACAGAGTTTTTGGCAAGTACAAATATATAGGCTGTAGATCTTGTATTAGGAG 840
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 catgtcgatgtactgtattgccaatgtcctgcctgattattgaaataatcagcaggac 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 841 TTAGTTCAATTTAAATGTTGTTGCCAACGCTGGCACCCTGACCCCGCATAGTCAAGTCAGGAC 900
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 agaataaaactgtggaatttttaa 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 901 AGACTTGTAAATATCAATGTTTGAA 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-08-202-186-18
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-18

Query Match      15.9%; Score 44.6; DB 1; Length 1105;
Best Local Similarity 55.4%; Pred. No. 1.2e-05;
Matches 108; Conservative 0; Mismatches 84; Indels 3; Gaps 1;

QY 58 aaatttaacccctcgatgtaccagtgtaatttagagtgatttaattatgccttgtaga 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 760 AATAGTTATATTGATATATCCAGATGCAAGAGAGGAATATTTAAACTATGTTTATTAGA 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 118 atgtgttaagaacagggcattcagttcggacaaatacgaaccccttagttatcttgggt 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 820 AGAATTTAAGAATGGAATTTTCAAGCGGGAATATGAACCCGTTTGAATG---T 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 cgaccatgtcgatgtactgtattgccaatgtcctgcctgattattgaaataatcagcag 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 877 AGAATATGTGGAAGTCATGTAATGCTAACTCTCTCCGAAGGAAGGAATCTTTTCAGA 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 ggacagaataaaact 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 937 AGATCGAATAAAGCT 951
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-08-202-186-16
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

Query Match      15.8%; Score 44.4; DB 1; Length 1103;
Best Local Similarity 55.8%; Pred. No. 1.4e-05;
Matches 106; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 63 taatcctcgatgtaccagtgtaatttagagtgatttaattatgccttgtagaagt 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 763 TTATATTTGATATTTCCAGATGCAAGAGGAATATTTAAACTATGTTTATTAGGAAT 822
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

QY	123	ttaaagaacagggttcagtcgagcaaatacagaccccttagttctcttggttgacc	182
Db	823	TTAAATGAATATTTCAAAGCGGGAATATGAACCGTTTTGAAATG---TAGAAT	879
QY	183	atdgcagtactcgatttgcgaatgtcgcgtcgtattattgaaatcagcaggaca	242
Db	880	ATGTGAAGTCATTGTAATGGCTAACTTCCTCCGAGGAAGGAATCTTTTCTGAAGATC	939
QY	243	gaataaaact	252
Db	940	GAATAAAGCT	949

```

RESULT      8
US-08-202-186-14
; Sequence 14, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:

```

[illegible]

```

Db      935 GATCGAATAAGTTG 949
      || ||||| ||
RESULT          9
US-08-202-186-12
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIRE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-12

```

```

Query Match      15.3%; Score 43; DB 1; Length 1111;
Best Local Similarity 54.9%; Pred. No. 4.2e-05;
Matches 107; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

QY 59 aatttaactctcgatgaccagagtgtaatttagagtgatttaaatatgcccctggttagaa 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 759 ATTGTTATATTTCATATTCCAAGATGCGAAGAGGATTATTAAATTATGGGTTATTAGAA 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 119 tctgttaagaacagggcattcagctcggcgcacaaatcacgaaccccttagttatctctggggttc 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 819 GAATTTAAGAATGGAATTAATTCAAAGCGGGAATAATGAACCCGCTTTTGAAGATAG--TA 875

QY 179 gaccatgcatctactctgatttgcacatgtcctgcctgatttattgaaatacagcagg 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 876 GAATATGTCGAAGTCATTGTAAATGGCTAACTTCCTCCGAAGGAAGGAATCTTTTCTGAA 935

QY 239 gacagaataaaactg 253
    || ||||| ||
Db 936 GATCGAATAAAGTTG 950

RESULT 10
US-08-202-186-17
; Sequence 17, Application US/08202186
; Patent No. 5756708
GENERAL INFORMATION:

```

```

; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-17

Query Match 15.2%; Score 42.8; DB 1; Length 1104;
Best Local Similarity 55.3%; Pred. No. 4.8e-05;
Matches 105; Conservative 0; Mismatches 82; Indels 3; Gaps 1;

QY 63 taatctcgatgaccaggtgtaattagagattttaaattatgcccctgttagaatgtg 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 764 TATATTTTCATATCCAGATGCAAGAGGAATATTTAAACTATGCTTTATTAGAAGAT 823

QY 123 ttaagacagggcattcagttcgacgaataacgaaccccttagttatcttgggttcgacc 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 824 TTAATAATGGAATTATTCAAAGCGGGAATATGAACCCGTTTGAAATTTG---TAGAAT 880

QY 183 atgtcgatgactcgatttgccaatgtctcctgctgatttattgaaatcagcagggaca 242
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 881 ATGTGAAGTCATGTGATGGTCTACTTCCCTCCGAAGGAATCTTTCTGAAGATC 940

QY 243 gaataaaact 252
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 941 GAATAAAGCT 950

RESULT 11
US-08-973-068-28
; Sequence 28, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068

```

```

; CURRENT FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; EARLIER FILING DATE: 1996-05-31
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 982
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; US-08-973-068-28

Query Match 14.7%; Score 41.4; DB 3; Length 982;
Best Local Similarity 54.4%; Pred. No. 0.00013;
Matches 106; Conservative 0; Mismatches 86; Indels 3; Gaps 1;

QY 59 aatttaactctcgatgtaccaggtgtaatttagagattttaaattatgcccctgttagaa 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 attgtattattgattatccaaagatgcaaaaggagattttaaattatgcccctgttagag 287

QY 119 tgtgttaagaacagggcatttcgacgaataacgaaccccttagttattcttgggttc 178
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 gaatttaagaatggaataattcaaacgggaaataatgaacccgttttgaagatag---ta 344

QY 179 gaccatgtcgatgactcgatttgccaatgtctcctgctgatttattgaaatcagcaggg 238
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 gaatatgtcgagtcattgttaattgcttaacttctcctcgaaggagaatcttttctgaa 404

QY 239 gacagaataaaactg 253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 gatcgaataaagtgtg 419

RESULT 12
US-08-202-186-13
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular

```


Search completed: October 24, 2001, 10:00:29
Job time: 214 sec

```

RESULT 15
US-08-202-186-10
; Sequence 10, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus

```

? APPLICANT: HARDING, ROBERT W.
 ? TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
 ? NUMBER OF SEQUENCES: 60
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Foley & Lardner
 ? STREET: 3000 K Street, N.W.
 ? CITY: Washington, D.C.
 ? COUNTRY: USA
 ? ZIP: 20007-5109
 ?

```

: ZIP: 20007-5109
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patent In Release #1.0, Version #1.25

```

```

; SOFTWARE.  FOR TECHNICAL ASSISTANCE, CONTACT THE
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/202,186

```

APPLICATION NUMBER: US/08/202,180
FILING DATE: 24-FEB-1994
CLASSIFICATION: 435

CLASSIFICATION: 433
ATTORNEY/AGENT INFORMATION:

NAME: JEFFERY DONALD D

NAME: JEFFERY, DONALD D.
REGISTRATION NUMBER: 19-980

REGISTRATION NUMBER: 19,300
REFERENCE/DOCKET NUMBER: 71611/102 FIKE

; REFERENCE/DOCREF NUMBER: 71
; TELECOMMUNICATION INFORMATION:

TELECOMMUNICATION INFORMATION
TELEPHONE: 202 672 5300

TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399

TELEFAX: 202 672 3399
TELE: 904136

TELEX: 904136
INFORMATION FOR SEQ ID NO: 10:

;; INFORMATION FOR SEQ ID NO: 1
:
: SEQUENCE CHARACTERISTICS:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs

```

LENGTH: III base pairs
TYPE: nucleic acid

TYPE: nucleic acid
STRANDNESS: single

STRANDEDNESS: single

TOPLOGY: circular

US-08-202-186-10

Query Match	14.2%	Score 40;	DB 1;	Length 1111;
Best Local Similarity	53.5%;	Pred. No. 0.00041;		
Matches 107:	Conservative	0:	Mismatches 90;	Indels 3; Gaps 1;

54 aacgaaatttaatcctcgatgtaccaggtgtaatattagaagtatttaaattatgccctgt 113

[illegible]

754 AACATATTCTTATAATTGATATTTCCAGATGCAAGAGGATTATTAAATTATGGGTAT 813

UD /54 AAGATAATGTTAATAATTGAATAATCCAAATAATGCACCAATGGCTATATTAATATATATCGCCTAAA

0.7 114 t a g a a t c t a t t a a g a a c a g a g c a t t c a a t t c g g a c a a a t a c g a a c c c c t t a g t t a t c t t g 173

QY 114 CAGACGCGCCAGAACGGGCACCAGCCCGACCAACACGCCCCC

814 TAGAGGAAATTAAAGAGTGGAAATAATCAAGCGGGAAATATGAACCCGTTTGAAGATAG 873

DD 814 TAGAGGAAATTAAAGAGTGGAAATAATCAGAGCGGAGAAATATGAAACCGCTATTAGAGGAAAG 578

174 atttccaccatctaccatctactctatttaccaatctcctgcctgattattgaaaaatca 233

I/4 ggttcgaccatgtgcattcccatccttcccgaaccca ccaaaaccc

Qy | | | | | | | | | | | | | | | | | | | |

24 674 - E A C A A E E C C A A C E C A T C T A A T C C C T A A C T T C C T C C A A G G A A T C T T T 930

DB 8/4 ---TAGAAATATGTCGAAGTCATTTGTAATGGCTAACITCCGAAAGGAAGGAATCTTTT 930

[illegible]

Qy 234 gcaggacagaataaaaactg 253

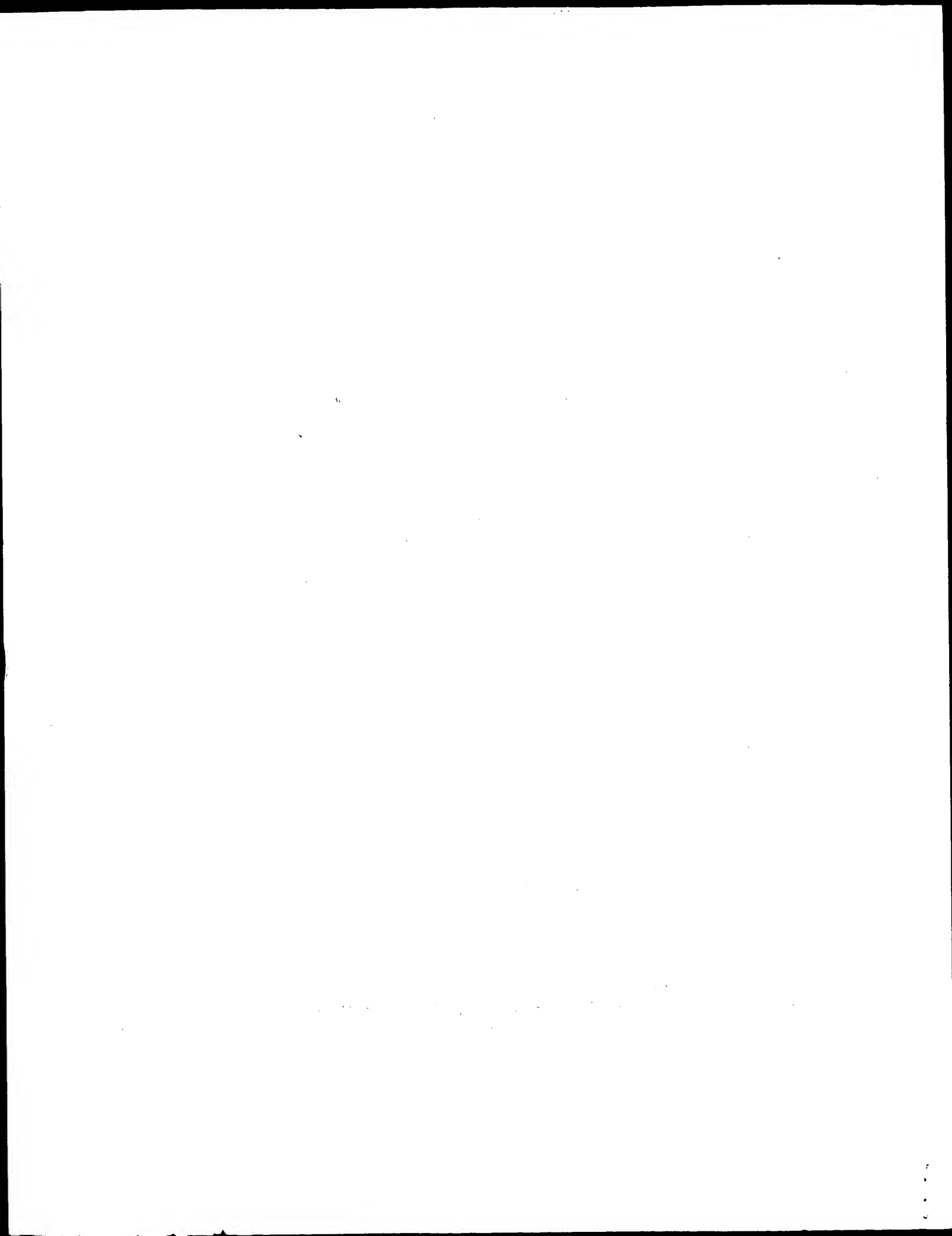
2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

Db 931 CTGAAGATCGAATAAAGTTG 950

Thu Oct 25 13:08:29 2001

us-09-462-955-1_copy_711_991.rni

Page 9



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:25 ; Search time 393.52 Seconds
(without alignments)
49.464 Million cell updates/sec

Title: US-09-462-955-2
Perfect score: 31
Sequence: 1 agccgcggggtaatactagccccgcgcgt 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*

1:	/SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	15	AAQ56057
2	31	100.0	31	20	AAQ56057
3	31	100.0	31	20	AAQ56057
C 4	19.8	63.9	31	20	AAQ56057
C 5	19.8	63.9	31	20	AAQ56057
C 6	19.8	63.9	31	20	AAQ56057
C 7	19.8	63.9	31	20	AAQ56057
C 8	18.4	59.4	415	21	AAQ56057
C 9	18.2	58.7	60	18	AAQ56057
C 10	18.2	58.7	60	18	AAQ56057
C 11	18.2	58.7	352	18	AAQ56057

C 12	18.2	58.7	533	18	AAQ56057
C 13	18.2	58.7	623	18	AAQ56057
C 14	18.2	58.7	630	20	AAQ56057
C 15	18.2	58.7	684	21	AAQ56057
C 16	18.2	58.7	1075	18	AAQ56057
C 17	18.2	58.7	1075	18	AAQ56057
C 18	18.2	58.7	1075	18	AAQ56057
C 19	18.2	58.7	1089	18	AAQ56057
C 20	18.2	58.7	1089	18	AAQ56057
C 21	18.2	58.7	1089	18	AAQ56057
C 22	18.2	58.7	1110	19	AAQ56057
C 23	17.8	57.4	389	21	AAQ56057
C 24	17.8	57.4	811	21	AAQ56057
C 25	17.8	57.4	1284	8	AAQ56057
C 26	17.8	57.4	1284	11	AAQ56057
C 27	17.8	57.4	1284	19	AAQ56057
C 28	17.8	57.4	2974	14	AAQ56057
C 29	17.4	56.1	1438	21	AAQ56057
C 30	17.2	55.5	1340	21	AAQ56057
C 31	17.2	55.5	2284	22	AAQ56057
C 32	17.2	55.5	2450	20	AAQ56057
C 33	17.2	55.5	4086	14	AAQ56057
C 34	17.2	55.5	12019	20	AAQ56057
C 35	17.2	55.5	437	21	AAQ56057
C 36	17.2	55.5	480	21	AAQ56057
C 37	17.2	55.5	20387	19	AAQ56057
C 38	17.2	55.5	26338	19	AAQ56057
C 39	17.2	55.5	96988	21	AAQ56057
C 40	17.2	55.5	117213	19	AAQ56057
C 41	16.8	54.2	477	20	AAQ56057
C 42	16.8	54.2	521	21	AAQ56057
C 43	16.8	54.2	1719	19	AAQ56057
C 44	16.8	54.2	4140	13	AAQ56057
C 45	16.8	54.2	8467	20	AAQ56057

ALIGNMENTS

RESULT 1
AAQ56057
ID AAO56057 standard; DNA; 31 BP.
XX
AC AAO56057;
XX
DT 12-AUG-1994 (first entry)
XX
DE Coconut Foliar Decay Virus promoter stem-loop.
XX
KW Coconut Foliar Decay Virus; CPDV; strong promoter; tissue-specific;
KW phloem-specific; stem-loop structure; transgenic plant; ds.
XX
OS Coconut Foliar Decay Virus.
XX
FH Key Location/Qualifiers
stem_loop 1..31
FT /*tag= a
FT /function= promoter
FT /note= "loop has homology to geminivirus sequence"
XX
PN DE4306832-C.
XX
PD 24-FEB-1994.
XX
PF 04-MAR-1993; 93DE-4306832.
XX
PR 04-MAR-1993; 93DE-4306832.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Becker D, Randles JW, Rohde W, Salamini F;
XX WPI; 1994-058406/08.
DR

Banana bunchy top
Banana bunchy top
Nucleotide sequenc
Human prostate can
Banana bunchy top
Banana bunchy top
Banana bunchy top
Banana bunchy top
Banana bunchy top
Banana bunchy top
N. meningitidis pa
Maize gibberellin
N-myc 1 proto-onco
N-myc 1 clone. A
nmyc gene fragment
Genomic DNA sequenc
Nucleotide sequenc
Wnt-4AF and Wnt-5c
Human MyoD1 gene p
Nucleotide sequenc
Human myoD gene an
Alcaligenes sp. Po
Aspergillus oryzae
Myrtaceae microsat
HSV-2 strain S85 C
HSV-2 strain S85 C
BAC containing rep
HSV-2 strain S85 C
Polynucleotide seq
Trichoderma reesei
Human par-4 gene.
Encodes acid alpha
Polynucleotide seq

XX Use of coconut foliar decay virus DNA as promoter - for
PT tissue-specific gene expression in transgenic plants
XX
XX Claim 1; Fig 2; 8pp; German.
XX A DNA fragment from the CFDV genome can be used as a phloem-specific
CC promoter in the construction of transgenic plants. The promoter is
CC strong; it has 30-50% of the activity of the CaMV 35S promoter in
CC tobacco plants.
XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 31; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagccccgcgcgt 31
|||||
DB 1 agccgcggggtaataactagccccgcgcgt 31

RESULT 2
AAx02688
ID AAX02688 standard; DNA; 31 BP.
XX
AC AAX02688;
XX
DT 10-MAY-1999 (first entry)
XX
DE Gemini virus DNA fragment stem loop.
XX
KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX
OS Gemini virus.
XX
FH Key Location/Qualifiers
FT stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX
PN DE19730502-A1.
XX
PD 21-JAN-1999.
XX
PF 16-JUL-1997; 97DE-1030502.
XX
PR 16-JUL-1997; 97DE-1030502.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX WPI; 1999-096863/09.
XX
XX Coconut foliar decay virus promoters - for gene expression in
XX plants, bacteria and yeasts
XX
XX Disclosure; Fig 2; 14pp; German.
XX
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
CC that includes the stem-loop structure of CFDV DNA but lacks the
CC translation start codons of open reading frames ORF1 and/or ORF2. The
CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
CC promoter for tissue-specific (especially phloem-specific) gene expression
CC in plants and for production of chimeric constructs for transient or
CC in plants and for production of chimeric constructs for transient or

CC stable expression. Certain fragments of CFDV DNA have stronger promoter
CC activity in E. coli than the CaMV 35S promoter.
XX
XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagccccgcgcgt 31
|||||
DB 1 agccgcggggtaataactagccccgcgcgt 31

RESULT 3
AAx02686
ID AAX02686 standard; DNA; 31 BP.
XX
AC AAX02686;
XX
DT 10-MAY-1999 (first entry)
XX
DE Gemini virus DNA fragment stem loop.
XX
KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX
OS Gemini virus.
XX
FH Key Location/Qualifiers
FT stem_loop 1..31 /*tag= a
FT misc_binding 1..11 /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31 /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX
PN DE19730535-A1.
XX
PD 21-JAN-1999.
XX
PF 16-JUL-1997; 97DE-1030535.
XX
PR 16-JUL-1997; 97DE-1030535.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX WPI; 1999-096867/09.
XX
XX Coconut foliar decay virus promoters - for gene expression in
XX plants, bacteria and yeasts
XX
XX Disclosure; Fig 2; 12pp; German.
XX
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
CC that includes the stem-loop structure of CFDV DNA but lacks the
CC translation start codons of open reading frames ORF1 and/or ORF2. The
CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
CC promoter for tissue-specific (especially phloem-specific) gene expression
CC in plants and for production of chimeric constructs for transient or
CC stable expression. Certain fragments of CFDV DNA have stronger promoter
CC activity in E. coli than the CaMV 35S promoter.
XX
XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;
SQ

Query Match 100.0%; Score 31; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagcccccgcgcgt 31
 |||||
 Db 1 agccgcggggtaataactagcccccgcgcgt 31

RESULT 4

AAQ56057/c
 ID AAQ56057 standard; DNA; 31 BP.

AC AAQ56057;

XX 12-AUG-1994 (first entry)

XX Coconut Foliar Decay Virus promoter stem-loop.

KW Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;
 KW phloem-specific; stem-loop structure; transgenic plant; ds.

XX Coconut Foliar Decay Virus.

FT Key Location/Qualifiers
 FT stem_loop 1..31

FT /*tag= a
 FT /function= promoter

FT /note= "loop has homology to geminivirus sequence"

XX DE4306832-C.

PD 24-FEB-1994.

XX 04-MAR-1993; 93DE-4306832.

XX 04-MAR-1993; 93DE-4306832.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Becker D, Randles JW, Rohde W, Salamini F;

XX WPI; 1994-058406/08.

XX Use of coconut foliar decay virus DNA as promoter - for
 XX tissue-specific gene expression in transgenic plants

XX Claim 1; Fig 2; 8pp; German.

XX A DNA fragment from the CFDV genome can be used as a phloem-specific
 CC promoter in the construction of transgenic plants. The promoter is
 CC strong; it has 30-50% of the activity of the CamV 35S promoter in
 CC tobacco plants.

XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 63.9%; Score 19.8; DB 15; Length 31;

Best Local Similarity 77.4%; Pred. No. 3.9;

Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagcccccgcgcgt 31
 |||||

Db 31 AGCCGCGGGGCTAGTATTACCCCGCGCT 1

RESULT 5

AAQ02688/c

ID AAX02688 standard; DNA; 31 BP.

XX AAX02688;

XX 10-MAY-1999 (first entry)

XX Gemini virus DNA fragment stem loop.

XX Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
 KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
 XX Gemini virus.

FT Key Location/Qualifiers

FT stem_loop 1..31

FT /*tag= a

FT misc_binding 1..11

FT /*tag= b

FT /note= "Region binds to nucleotides 21 to 31"

FT misc_binding 21..31

FT /*tag= c

FT /note= "Region binds to nucleotides 1 to 11"

XX DE19730502-A1.

XX 21-JAN-1999.

XX 16-JUL-1997; 97DE-1030502.

XX 16-JUL-1997; 97DE-1030502.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;

XX WPI; 1999-096863/09.

XX Coconut foliar decay virus promoters - for gene expression in
 XX bacteria and yeasts

XX Disclosure; Fig 2; 14pp; German.

XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
 CC that includes the stem-loop structure of CFDV DNA but lacks the
 CC translation start codons of open reading frames ORF1 and/or ORF2. The
 CC new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
 CC promoter for tissue-specific (especially phloem-specific) gene expression
 CC in plants and for production of chimeric constructs for transient or
 CC stable expression. Certain fragments of CFDV DNA have stronger promoter
 CC activity in *E. coli* than the CamV 35S promoter.

XX Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 63.9%; Score 19.8; DB 20; Length 31;

Best Local Similarity 77.4%; Pred. No. 3.9;

Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcggggtaataactagcccccgcgcgt 31
 |||||

Db 31 AGCCGCGGGGCTAGTATTACCCCGCGCT 1

RESULT 6

AAQ02686/c

ID AAX02686 standard; DNA; 31 BP.

XX AAX02686;

XX 10-MAY-1999 (first entry)

XX Gemini virus DNA fragment stem loop.

XX Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;

KW yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.

XX Gemini virus.

XX Key Location/Qualifiers

FT stem_loop 1..31

FT misc_binding /*tag= a.
1..11.
/*tag= b
FT /note= "Region binds to nucleotides 21 to 31"
FT 21..31
FT /tag= c
FT /note= "Region binds to nucleotides 1 to 11"
XX
XX DE19730535-A1.
XX
XX 21-JAN-1999.
XX
XX PF 16-JUL-1997; 97DE-1030535.
XX
XX PR 16-JUL-1997; 97DE-1030535.
XX
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX PI Becker D, Hehn A, Randles JW, Rohde W, Salami F;
XX
XX WPI; 1999-096867/09.
XX
XX Coconut foliar decay virus promoters - for gene expression in
XX plants, bacteria and yeasts
XX
XX PS Disclosure; Fig 2; 12pp; German.
XX
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX that includes the stem-loop structure of CFDV DNA but lacks the
XX translation start codons of open reading frames ORF1 and/or ORF2. The
XX new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX promoter for tissue-specific (especially phloem-specific) gene expression
XX in plants and for production of chimeric constructs for transient or
XX stable expression. Certain fragments of CFDV DNA have stronger promoter
XX activity in E. coli than the CAMV 35S promoter.
XX
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 63.9%; Score 19.8; DB 20; Length 31;
Best Local Similarity 77.4%; Pred. NO. 3.9;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgct 31
|||||
DB 31 AGCCGCGGGGCTAGTATTACCCCGCGGCT 1

RESULT 7
AAV27112
ID AAV27112 standard; DNA; 35099 BP.
XX
XX AC AAV27112;
XX
XX DT 28-SEP-1998 (first entry)
XX
XX DE Adenovirus 17.
XX
XX KW Adeno virus 17; Ad17; vector; gene therapy; cystic fibrosis;
XX alpha-antitrypsin deficiency; respiratory disease; ss.
XX
XX OS Mastadenovirus 17.
XX
XX PN WO9822609-A1.
XX
XX PD 28-MAY-1998.
XX
XX PF 20-NOV-1997; 97WO-US21494.
XX
XX PR 20-NOV-1996; 96US-0752760.
XX
XX PA (GENZ) GENZYME CORP.
XX

PI Armentano DE, Gregory RJ, Smith AE;
XX
XX WPI; 1998-312493/27.
XX
XX Recombinant adenovirus type 2 vector - useful for targetting
XX biologically active proteins, used to, e.g. treat cystic fibrosis
XX
XX PS Disclosure; Page 28-38; 67pp; English.
XX
XX This is the complete nucleotide (nt) sequence of adenovirus 17
XX (Ad17). A claimed chimeric adenoviral vector comprises a nt
XX sequence of an adenovirus in which all or part of a gene encoding
XX a protein facilitating attachment to, or internalisation into, a
XX target mammalian cell is replaced by all or part of the
XX corresponding gene for a second adenovirus from subgroup D,
XX preferably selected from Ad9, Ad15, Ad17, Ad19, Ad20, Ad22, Ad26,
XX Ad27, Ad28, Ad30 and Ad39. The vector further comprises a
XX transgene operably linked to a eukaryotic promoter to allow
XX expression in a mammalian cell. The replaced gene preferably
XX encodes an Ad fibre such as Ad17 fibre (see AAV27113) and/or an Ad
XX penton base such as Ad17 penton base (see AAV27114). The chimeric
XX vector is used to target biologically active proteins to airway
XX epithelial cells, especially for gene therapy of diseases such as
XX cystic fibrosis or alpha-antitrypsin deficiency. The vectors are
XX based on the discovery that proteins from the subgroup D viruses,
XX a group not normally associated with human respiratory diseases,
XX can effectively bind and internalise within human airway
XX epithelial cells.
XX
XX SQ Sequence 35099 BP; 7988 A; 9877 C; 9978 G; 7252 T; 4 other;

Query Match 60.6%; Score 18.8; DB 19; Length 35099;
Best Local Similarity 76.7%; Pred. NO. 22;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcggc 30
|||||
DB 15104 accgctgggtattactagggccagcgc 15133

RESULT 8
AAV27112
ID AAF09348 standard; cDNA; 415 BP.
XX
XX AC AAF09348;
XX
XX DT 13-MAR-2001 (first entry)
XX
XX DE Fusarium venenatum EST SEQ ID NO:1871.
XX
XX KW Multiple gene expression; filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX culture condition; environmental stress; spore morphogenesis;
XX metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX OS Fusarium venenatum.
XX
XX PN WO2000056762-A2.
XX
XX PD 28-SEP-2000.
XX
XX PF 22-MAR-2000; 2000WO-US07781.
XX
XX PR 22-MAR-1999; 99US-0273623.
XX
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
XX
XX PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX

Monitoring differential expression of genes in filamentous fungal cells
uses fluorescence-labeled nucleic acids isolated from the cells and a
substrate of expressed sequence tags -

Claim 86; Page 1082; 3161pp; English.

The present invention describes a method for monitoring differential
expression of genes in a first filamentous fungal (FF) cell relative to
expression of the same genes in one or more second filamentous fungal
cells. The method uses fluorescence-labeled nucleic acids isolated from
the FF cells and a substrate of expressed sequence tags (EST). The ESTs
are used in the methods for monitoring differential expression of genes
in a first filamentous fungal (FF) cell relative to expression of the
same genes in one or more second filamentous fungal cells. Monitoring
the global expression of genomes from FF cells allows the production
potential of the microorganisms to be improved. New genes may be
discovered, possible functions of unknown open reading frames can be
identified, and gene copy number variation and stability can be
monitored. The expression of genes can be used to study how FF cells
adapt to changes in culture conditions, environmental stress, spore
morphogenesis, recombination, metabolic or catabolic pathway
engineering. Using ESTs provides several advantages over genomic or
random cDNA clones including elimination of redundancy as one spot on an
array equals one gene or open reading frame, and organisation of the
microarrays based on function of the gene products to facilitate
analysis of the results. AAF07478 to AAF11247 represents ESTs from
Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from *Aspergillus*
niger; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
all specifically claimed in the present invention.

Sequence 415 BP; 95 Å; 122 C; 113 G; 79 T; 6 other:

	Query Match	59.4%	Score 18.4;	DB 21;	Length 415;
	Best Local Similarity	78.6%	Pred. No. 21;		
	Matches	22;	Conservative	0;	Mismatches 6;
					Indels 0;
					Gaps 0;
Qy	2	gccacgggggtaactagccgcggg	29		
Db	38	gccacgagagtactacgacgcgcggg	65		

RESULT	9
ID	NAT49396/c
DD	AAT49396 standard; DNA; 60 BP.
XX	AC
XX	AAT49396;
DT	23-AUG-1997 (first entry)
XX	Banana bunchy top virus DNA component 6 stem-loop common region.
DE	BTV; intergenic region; promoter; transgenic plant; ss.
XX	Banana bunchy top virus.
SS	
HH	Key
T	stem_loop
T	Location/Qualifiers
T	17..47
T	/tag= a
T	misc_binding
T	17..26
T	/tag= b
T	/note= "binds to nts 47..56"
T	misc_structure
T	27..37
T	/tag= c
T	/note= "loop"
T	misc_binding
T	38..47
T	/tag= d
T	/note= "binds to nts 26..35"
X	WO9638554-A1.
X	

[illegible]

PF 31-MAY-1996; 96WO-AU00335.
 XX
 XX 31-MAY-1995; 95AU-0003285.
 PR
 XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 PA
 XX Becker DK, Beetham PR, Dale JL, Dugdale B, Hafner GJ;
 PI Harding RM;
 XX WPI; 1997-034368/03.
 DR
 XX DNA from intergenic region of banana bunchy top virus DNA component
 XX - useful for promoting, enhancing, regulating or modifying
 PT transcription of a non-BBTV gene in transgenic plants
 PT
 XX
 PS Claim 4; Fig 3; 80pp; English.
 PS
 XX Stem-loop common regions (AAT49391-96) are conserved in DNA
 CC components 1-6 (see also AAT49386-90) of banana bunchy top virus
 CC virus (BBTV). Each component has an 11-nucleotide loop sequence
 CC of which 9 consecutive nucleotides are conserved between all 6
 CC components, and a 10 bp stem sequence of which 14 nucleotides are
 CC conserved. Intergenic regions (see also AAT49399-409) of components
 CC 1-6, including regions contg. these stem-loop common regions, are
 CC useful for promoting, enhancing, regulating or modifying
 CC transcription of non-BBTV genes in monocotyledon or dicotyledon
 CC transgenic plants.
 CC
 XX Sequence 69 BP; 14 A; 20 C; 21 G; 14 T; 0 other;
 SQ

Query Match 58.7%; Score 18.2; DB 18; Length 69;
 Best Local Similarity 74.2%; Pred. No. 22;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgcgt 31
 ||| ||||| ||| ||| |||
 Db 56 AGCGTGGGGTAATAATAGTCCCGCGCT 26

RESULT 11
 AAT49407/c
 ID AAT49407 standard; DNA; 352 BP.
 AC AAT49407;
 XX
 XX 23-AUG-1997 (first entry)
 DT
 XX Banana bunchy top virus DNA intergenic region 6 insert in pBT6.2.
 DE
 XX BBTV; intergenic region; promoter; transgenic plant; ds.
 KW
 XX Banana bunchy top virus.
 OS
 XX WO9638554-A1.
 PN
 XX 05-DEC-1996.
 PD
 XX 31-MAY-1996; 96WO-AU00335.
 PF
 XX 31-MAY-1995; 95AU-0003285.
 PR
 XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 PA
 XX Becker DK, Beetham PR, Dale JL, Dugdale B, Hafner GJ;
 PI Harding RM;
 XX WPI; 1997-034368/03.
 DR
 XX DNA from intergenic region of banana bunchy top virus DNA component
 XX - useful for promoting, enhancing, regulating or modifying
 PT transcription of a non-BBTV gene in transgenic plants
 PT
 XX

PS Claim 4; Fig 13b; 80pp; English.
 XX
 XX Intergenic regions (AAT49399-409) are derived from genomic DNA
 CC components 1-6 (see also AAT49386-90) of banana bunchy top virus
 CC (BBTV). A component 6 intergenic region insert (AAT49407) in pBT6.2
 CC was obtd. by deletion of 272 bp from the 5' end of the component
 CC 6 intergenic region given in AAT49402 and cloning of the fragment
 CC into pBI101.3. The full intergenic region of component 6 has
 CC promoter activity comparable to that of the 800 bp CamV 35S
 CC promoter from pBI121. The 272 bp deletion caused a significant
 CC increase in promoter activity which was maintained with a further
 CC 112 bp 5' deletion (see also AAT49408), but promoter activity was
 CC significantly lost with a further 75 bp deletion (AAT49409).
 XX
 XX Sequence 352 BP; 105 A; 70 C; 89 G; 88 T; 0 other;
 SQ

Query Match 58.7%; Score 18.2; DB 18; Length 352;
 Best Local Similarity 74.2%; Pred. No. 26;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgcgt 31
 ||| ||||| ||| ||| |||
 Db 104 AGCAGGGGGTAATAATAGTCCCGCGTGT 74

RESULT 12
 AAT49401/c
 ID AAT49401 standard; DNA; 533 BP.
 AC AAT49401;
 XX
 XX 23-AUG-1997 (first entry)
 DT
 XX Banana bunchy top virus DNA intergenic region 3 insert in pBT3.1.
 DE
 XX BBTV; intergenic region; promoter; transgenic plant; ds.
 KW
 XX Banana bunchy top virus.
 OS
 XX Key Location/Qualifiers
 FT misc_binding 329..337
 FT /*tag= a
 FT /note= "Adh1 US1 motif"
 FT
 XX WO9638554-A1.
 PN
 XX 05-DEC-1996.
 PD
 XX 31-MAY-1996; 96WO-AU00335.
 PF
 XX 31-MAY-1995; 95AU-0003285.
 PR
 XX (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
 PA
 XX Becker DK, Beetham PR, Dale JL, Dugdale B, Hafner GJ;
 PI Harding RM;
 XX WPI; 1997-034368/03.
 DR
 XX DNA from intergenic region of banana bunchy top virus DNA component
 XX - useful for promoting, enhancing, regulating or modifying
 PT transcription of a non-BBTV gene in transgenic plants
 PT
 XX
 PS Claim 4; Fig 11; 80pp; English.
 PS
 XX Intergenic regions (AAT49399-409) are derived from genomic DNA
 CC components 1-6 (see also AAT49386-90) of banana bunchy top virus
 CC (BBTV). Intergenic region bbtvpro3 (AAT49401) comprises a BBTV
 CC component 3 insert in pBT3.1, a plasmid obtd. by PCR amplification
 CC (see also AAT49419-20) of BBTV diseased banana nucleic acid extract,
 CC subcloning of amplified intergenic region 3 into pGEM3zf+ vector
 CC and further cloning as a HindIII-BamHI fragment into pBI101.3.
 CC

Db 203 AGCGGGGAGAACTACAGTCCCGCGGCT 173

Best Local Similarity 74.2%; Pred. No. 27;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Search completed: October 24, 2001, 10:07:26
Job time: 631 sec

OY 1 agccgagggttaataactagcccccggt 31
||| ||||| ||| ||| |||
DB 378 AGCAGCGGGGTAAATATAGTCCCGGTCT 348

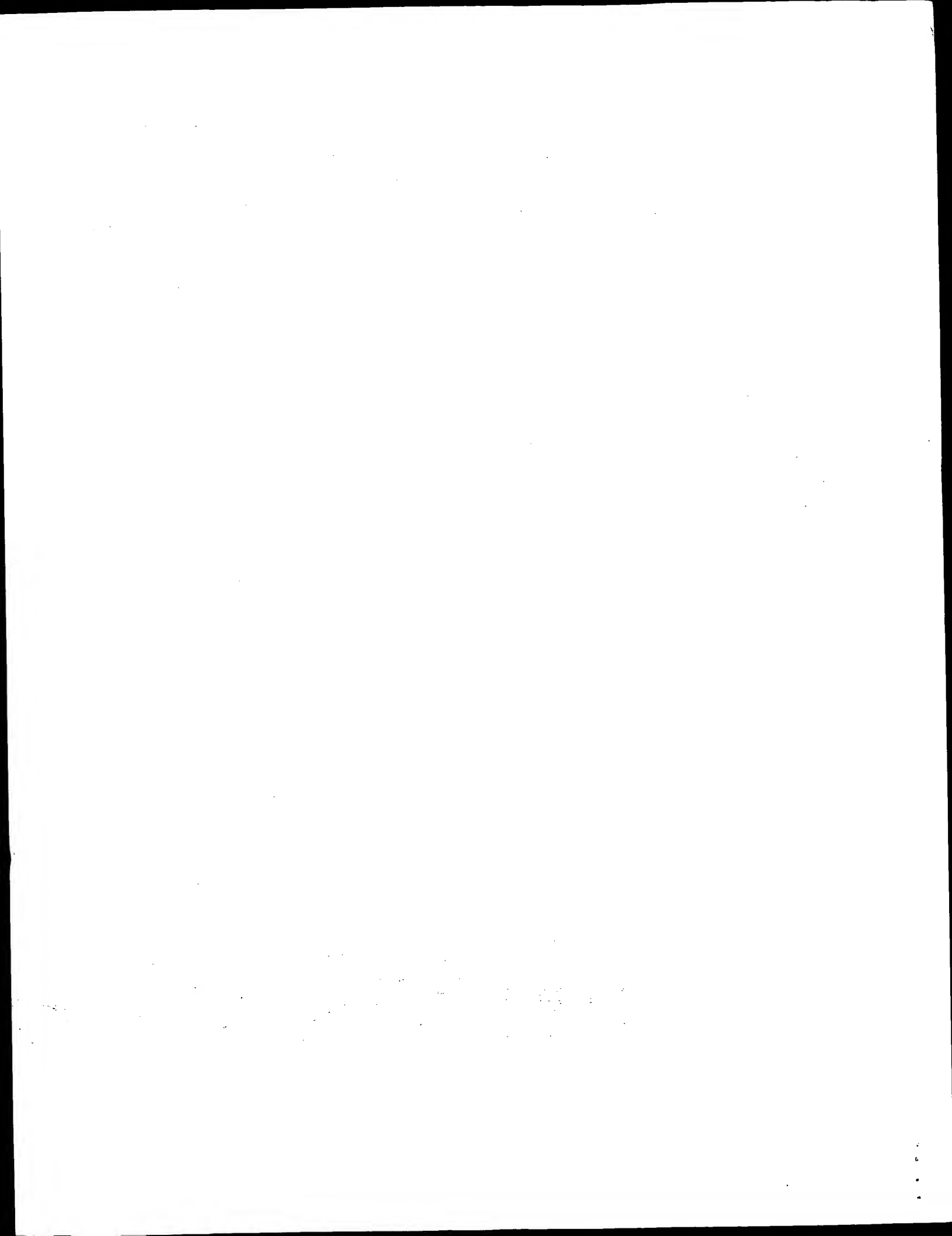
RESULT 15
AAFI6045/c
ID AAFI6045 standard; cDNA; 684 BP.
XX
AC AAFI6045;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:480.
XX
DE Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulvular; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.

XX WO200055174-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05988.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM;
PI
XX WPI; 2000-587513/55.
DR P-PSDB; AAB56842.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer
XX
XX Claim 1; Page 952; 2338pp; English.

XX AAFI5566 to AAFI6505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulvular, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAFI6506 to AAFI6514 to
CC AAB57303 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 684 BP; 111 A; 208 C; 248 G; 112 T; 5 other;

Query Match 58.7%; Score 18.2; DB 21; Length 684;
Best Local Similarity 74.2%; Pred. No. 28;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 agccgagggttaataactagcccccggt 31
||| ||||| ||| ||| |||



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:48 ; Search time 5479.82 Seconds
(without alignments)
53.476 Million cell updates/sec

Title: US-09-462-955-2

Perfect score: 31

Sequence: 1 agccgcgggggtaataactagccccgcggct 31

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
10: gb_est10: *
11: gb_est11: *
12: gb_est12: *
13: gb_est13: *
14: gb_est14: *
15: gb_est15: *
16: gb_est16: *
17: gb_est17: *
18: gb_est18: *
19: gb_est19: *
20: gb_est20: *
21: gb_est21: *
22: gb_est22: *
23: gb_est23: *
24: gb_est24: *
25: gb_est25: *
26: gb_est26: *
27: gb_est27: *
28: gb_est28: *
29: gb_est29: *
30: gb_est30: *
31: gb_est31: *
32: gb_est32: *
33: gb_est33: *
34: gb_est34: *
35: gb_est35: *
36: gb_est36: *
37: gb_est37: *
38: gb_est38: *
39: gb_est39: *
40: gb_est40: *
41: gb_est41: *
42: gb_est42: *
43: gb_est43: *
44: gb_est44: *
45: gb_est45: *
46: gb_est46: *
47: gb_est47: *

44: em_esthum10: *
45: em_esthum11: *
46: em_esthum12: *
47: em_esthum13: *
48: em_esthum14: *
49: em_esthum15: *
50: em_esthum16: *
51: em_esthum17: *
52: em_esthum18: *
53: em_esthum19: *
54: em_esthum20: *
55: em_esthum21: *
56: em_esthum22: *
57: em_esthum23: *
58: em_esthum24: *
59: em_esthum25: *
60: em_esthum26: *
61: em_esthum27: *
62: em_esthum28: *
63: em_estin1: *
64: em_estin2: *
65: em_estin3: *
66: em_estin4: *
67: em_estin5: *
68: em_estomi: *
69: em_estom2: *
70: em_estov1: *
71: em_estov2: *
72: em_estpl1: *
73: em_estpl2: *
74: em_estpl3: *
75: em_estpl4: *
76: em_estpl5: *
77: em_estpl6: *
78: em_estpl7: *
79: em_estpl8: *
80: em_estpl9: *
81: em_estpl10: *
82: em_estrol: *
83: em_estrol2: *
84: em_estrol3: *
85: em_estrol4: *
86: em_estrol5: *
87: em_estrol6: *
88: em_estrol7: *
89: em_estrol8: *
90: em_estrol9: *
91: em_estrol10: *
92: em_estrol11: *
93: em_estrol12: *
94: em_estrol13: *
95: em_estrol14: *
96: em_estrol15: *
97: em_estrol16: *
98: em_estrol17: *
99: em_estrol18: *
100: em_estrol19: *
101: em_estrol20: *
102: gb_est25: *
103: gb_est26: *
104: gb_est27: *
105: gb_est28: *
106: gb_est29: *
107: gb_est30: *
108: gb_est31: *
109: gb_est32: *
110: gb_est33: *
111: gb_est34: *
112: gb_est35: *
113: gb_est36: *
114: gb_est37: *
115: gb_est38: *
116: gb_est39: *

us-09-462-955-2.rst

Thu Oct 25 13:08:43 2001

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: em_esthum29:*
169: em_esthum30:*
170: em_esthum31:*
171: em_esthum32:*
172: em_esthum33:*
173: em_estom3:*
174: em_estpl1:*
175: em_estro21:*
176: em_estro22:*
177: em_estro23:*
178: em_htc:*
179: gb_est107:*
180: gb_est108:*
181: gb_est109:*
182: gb_est110:*
183: gb_est111:*
184: em_gss_fun:*
185: em_gss_hum1:*
186: em_gss_hum2:*
187: em_gss_hum3:*
188: em_gss_hum4:*
189: em_gss_hum5:*
190: em_gss_hum6:*
191: em_gss_hum7:*
192: em_gss_hum8:*
193: em_gss_hum9:*
194: em_gss_inv1:*
195: em_gss_inv2:*
196: em_gss_inv3:*
197: em_gss_other:*
198: em_gss_pln1:*
199: em_gss_pln2:*
200: em_gss_pro:*
201: em_gss_rod1:*
202: em_gss_rod2:*
203: em_gss_rod3:*
204: em_gss_rod4:*
205: em_gss_rod5:*
206: em_gss_vrt1:*
207: em_gss_vrt2:*
208: em_gss_vrt3:*
209: gb_gss1:*
210: gb_gss2:*
211: gb_gss3:*
212: gb_gss4:*
213: gb_gss5:*
214: gb_gss6:*
215: gb_gss7:*
216: gb_gss8:*
217: gb_gss9:*
218: gb_gss10:*
219: gb_gss11:*
220: gb_gss12:*
221: gb_gss13:*
222: gb_gss14:*
223: gb_gss15:*
224: gb_gss16:*
225: gb_gss17:*
226: gb_gss18:*
227: gb_gss19:*
228: gb_gss20:*
229: gb_gss21:*
230: gb_gss22:*
231: gb_gss23:*
232: gb_gss24:*
233: gb_gss25:*
234: gb_gss26:*
235: gb_gss27:*
236: gb_gss28:*
237: gb_gss29:*
238: gb_gss30:*
239: gb_gss31:*
240: gb_gss32:*
241: gb_gss33:*
242: gb_gss34:*
243: em_gss_inv4:*
244: em_gss_rod6:*
245: em_gss_rod7:*
246: em_gss_rod8:*
247: gb_gss35:*
248: gb_gss36:*
249: gb_gss37:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21.4	69.0	905	220	CNS03052	AL221807 Tetraodon
2	21.4	69.0	1072	220	CNS02GRK	AL196697 Tetraodon
3	20.4	65.8	876	142	BE968958	BE968958 601649921
4	20.4	64.5	1843	144	BF128533	BF128533 601810627
5	19.8	63.9	865	141	BE902577	BE902577 601677225
6	19.8	63.9	890	168	BF695233	BF695233 602080459
7	19.6	63.2	1287	120	AW731193	AW731193 CA_Ea001
8	19.4	62.6	629	32	AV722569	AV722569 AV722569
9	19.4	62.6	714	106	AU006051	AU006051 AU006051
10	19.4	62.6	797	153	BG401207	BG401207 602465296
11	19.4	62.6	836	144	BF126760	BF126760 601650551
12	19.4	62.6	897	146	BF300720	BF300720 602031903
13	19.4	62.6	983	221	CNS04KIP	AL294874 Tetraodon
14	19.4	61.3	627	165	BE275297	BE275297 601122139
15	19.4	61.3	705	143	BF032195	BF032195 601452609
16	19.4	61.3	715	235	AQ938642	AQ938642 NLI-C08R
17	19.4	61.3	734	175	BG281907	BG281907 602402672
18	19.4	61.3	824	168	BF696175	BF696175 602124688
19	19.4	61.3	902	175	BG281858	BG281858 602403088
20	19.4	61.3	958	172	BF982728	BF982728 602304933
21	19.4	61.3	1160	172	BF971171	BF971171 602270763
22	18.8	60.6	501	155	BG561147	BG561147 EtESted79
23	18.8	60.6	537	19	AI352913	AI352913 MB73-4B P
24	18.8	60.6	909	142	BE959324	BE959324 601654233
25	18.8	60.6	912	146	BF263427	BF263427 HV_CEA000
26	18.8	60.6	944	140	BE779793	BE779793 601465480
27	18.8	60.6	971	222	CNS04Y22	AL312452 Tetraodon
28	18.8	60.6	1030	137	BF562027	BF562027 601345668
29	18.8	60.6	1072	172	BG030208	BG030208 602297502
30	18.6	60.0	431	155	BE557433	BE557433 EMI_44_A1
31	18.6	60.0	760	251	AZ900170	AZ900170 RPT-24-1
32	18.6	60.0	930	105	AL520966	AL520966 AL520966
33	18.6	60.0	954	138	BE620637	BE620637 601483609
34	18.4	59.4	115	166	BE348526	BE348526 ht7le09.x
35	18.4	59.4	224	104	A1969538	A1969538 wz68c09.x
36	18.4	59.4	277	161	BE577683	BE577683 BS577683
37	18.4	59.4	284	117	AW548262	AW548262 L003G11-
38	18.4	59.4	319	113	AW204439	AW204439 UI-H-B11-
39	18.4	59.4	321	143	BF058466	BF058466 7K31a07.x
40	18.4	59.4	340	229	AQ473306	AQ473306 C1PBI-El-
41	18.4	59.4	361	108	AU171662	AU171662 AU171662
42	18.4	59.4	369	21	A1523972	A1523972 tg98h02.x
43	18.4	59.4	401	108	AU170664	AU170664 AU170664
44	18.4	59.4	447	234	AQ853114	AQ853114 LMAJFV1_1
45	18.4	59.4	466	138	BE676163	BE676163 7f24b10.x

ALIGNMENTS

RESULT 1	
CNS03052	
LOCUS	905 bp DNA GSS 15-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 183N12 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL221807
VERSION	AL221807.1 GI:7880626
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 905)
AUTHORS	Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,

Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.	
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis	
Unpublished	
2 (bases 1 to 905)	
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.	
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	
Unpublished	
3 (bases 1 to 905)	
Genoscope.	
Direct Submission	
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases	
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.	
Location/Qualifiers	
1..905	
/organism="Tetraodon nigroviridis"	
/db_xref="taxon:99883"	
/clone="183N12"	
/clone_lib="G"	
/note="Genoscope sequence ID : C0AG183DG06LPI-end : T7"	
BASE COUNT	138 a 299 c 282 g 177 t 9 Others
ORIGIN	
Query Match	69.0%; Score 21.4; DB 220; Length 905;
Best Local Similarity	80.6%; Pred. No. 24;
Matches	25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY	1 agccgcggggtaactagcccccggcgt 31
Db	482 AGCCGCAGGGCTACTACCGCGCCCCCGCGCT 512
RESULT 2	
CNS02GRK	
LOCUS	1072 bp DNA GSS 13-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 13809 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL196697
VERSION	AL196697.1 GI:7834847
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Tetraodontidae; Tetraodon.
REFERENCE	1 (bases 1 to 1072)
AUTHORS	Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1072)
AUTHORS	Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 1072)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

10

BE902577
 VERSION BE902577.1 GI:10392908
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 865)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LCM841 row: b column: 08
 High quality sequence stop: 675.
 Location/Qualifiers
 1..865
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3959719"
 /clone_lib="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 204 a 225 c 287 g 149 t
 ORIGIN
 Query Match 63.9%; Score 19.8; DB 141; Length 865;
 Best Local Similarity 77.4%; Pred. No. 1.2e-02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 agccgcggggtaatactagccccgcgcgt 31
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 727 AGCCCGGGGTATATCTGGGCCCTGCGCT 757
 RESULT 6
 LOCUS BF695233 890 bp mRNA EST 22-DEC-2000
 DEFINITION 602080459f1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4244751 5', mRNA sequence.
 ACCESSION BF695233
 VERSION BF695233.1 GI:11980641
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 890)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM1057 row: f column: 16
 High quality sequence stop: 334.
 Location/Qualifiers
 1..890
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4244751"
 /clone_lib="NIH_MGC_81"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccattagcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCCATATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGCGATG-dr(30)EN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
 BASE COUNT 308 a 253 c 155 g 174 t
 ORIGIN
 Query Match 63.9%; Score 19.8; DB 168; Length 890;
 Best Local Similarity 77.4%; Pred. No. 1.2e-02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 agccgcggggtaatactagccccgcgcgt 31
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 685 AGCCACGGGCTAATAAGGCGCGCGCGCT 715
 RESULT 7
 LOCUS AW731193 1287 bp mRNA EST 08-MAR-2001
 DEFINITION GA_Ea0010K01 Gossypium arboreum 7-10 dpa fiber library Gossypium
 ACCESSION AW731193
 VERSION AW731193.1 GI:7628851
 KEYWORDS EST.
 SOURCE Gossypium arboreum.
 ORGANISM Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 1287)
 AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
 JOURNAL Unpublished (2000)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 High quality sequence stop: 1287.
 Location/Qualifiers
 1..1287
 /organism="Gossypium arboreum"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0010K01"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
 /lab_host="E. coli"

```

/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      234 a 382 c 450 g 214 t 7 others
ORIGIN

Query Match      63.2%; Score 19.6; DB 120; Length 1287;
Best Local Similarity 84.6%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 gccggggglaactagccccgcgcgc 30
Db 406 GCGGGGTAGTACTCGCCGCCGCCGC 431

RESULT 8
AV722569 629 bp mRNA EST 16-OCT-2000
LOCUS AV722569 HTB Homo sapiens cDNA clone HTBANE01 5', mRNA sequence.
DEFINITION AV722569
ACCESSION AV722569.1 GI:10825189
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 629)
AUTHORS Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z., and Han, Z.
TITLE Homo sapiens cDNA HTB clones
JOURNAL Unpublished (2000)
CONTACT Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzeg@hgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..629
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTBANE01"
/clone_lib="HTB"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 185 a 124 c 129 g 190 t 1 others
ORIGIN

Query Match      62.6%; Score 19.4; DB 32; Length 629;
Best Local Similarity 79.3%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 gccgggggtaatactagccccgcgcgc 30
Db 597 CCGCGTGGCGTAATAGCAGCCGCCGCTGC 625

RESULT 9
AU006051/c 714 bp mRNA EST 19-JAN-1999
LOCUS AU006051 Bombyx mori p50(Daizo) Bombyx mori cDNA clone ww40411,
DEFINITION AU006051
ACCESSION AU006051
VERSION AU006051.1 GI:4163435
KEYWORDS EST.
SOURCE domestic silkworm.

```

```

Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 714)
AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Establishment of cDNA database of Bombyx mori
JOURNAL Unpublished (1999)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
PROJECT = 'CREST project by JST'.
FEATURES
source
Location/Qualifiers
1..714
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/clone="ww40411"
/clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT 173 a 183 c 191 g 166 t 1 others
ORIGIN

Query Match      62.6%; Score 19.4; DB 106; Length 714;
Best Local Similarity 79.3%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagccccgcgcgc 29
Db 398 AGCCGCGGTAGTAGTCGCGCTGCGG 370

RESULT 10
BG401207/7c 797 bp mRNA EST 12-MAR-2001
LOCUS BG401207 NTH_MGC_75 Homo sapiens cDNA clone IMAGE:4593646 5',
DEFINITION BG401207
ACCESSION BG401207.1 GI:13294655
VERSION EST.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 797)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1333 row: o column: 23
High quality sequence stop: 586.
FEATURES
source
Location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4593646"
/clone_lib="NTH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCGCCGACATG-dT(30)BN-3' (where B = A,

```

REFERENCE
AUTHORS
1 (bases 1 to 983)
Roest-crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and

```
source
1. .bz/
/organism="Homo sapiens"
/bi_xref="taxon:9606"
/clone="IMAGE:3346377"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
```

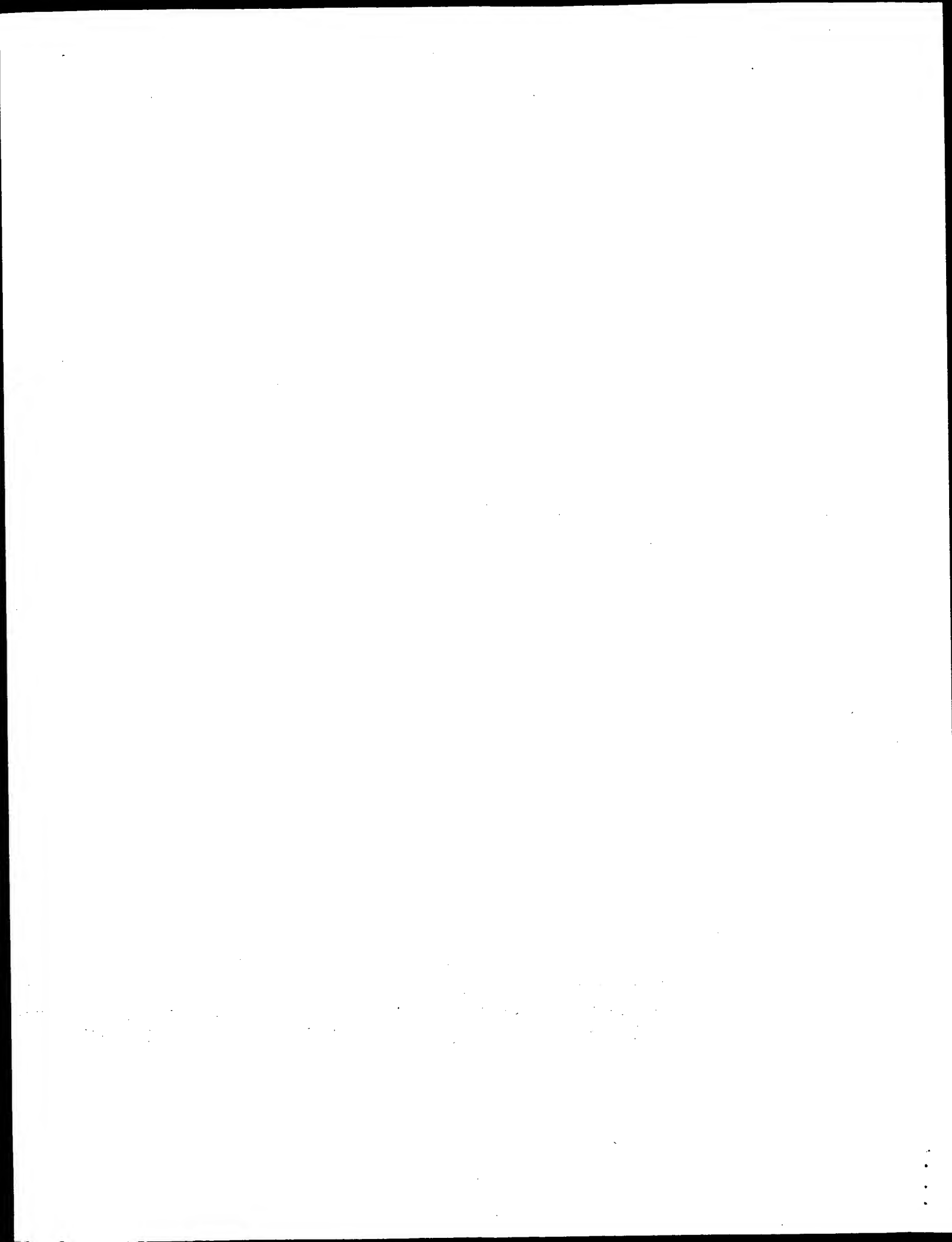
Qy 3 ccgcgggggtaataactagcccccg 29
|||||
pb 679 ccgcgggggggaaactatgcccccg 705
|||||

Thu Oct 25 13:08:43 2001

us-09-462-955-2.rst

Page 9

Search completed: October 24, 2001, 13:13:52
Job time: 11817 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw.model

Run on: October 24, 2001, 10:00:29 ; Search time 180.6 seconds
(without alignments)
32.495 Million cell updates/sec

Title: US-09-462-955-2

Perfect score: 31

Sequence: 1 agccgcggggtaatactagcccccgcgcgt 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	18.8	60.6	35081	2	US-08-752-760A-1
2	18.2	58.7	60	3	US-08-973-068-15
3	18.2	58.7	69	3	US-08-973-068-12
4	18.2	58.7	352	3	US-08-973-068-30
5	18.2	58.7	547	3	US-08-973-068-24
6	18.2	58.7	622	3	US-08-973-068-29
7	18.2	58.7	624	3	US-08-973-068-27
8	18.2	58.7	1075	1	US-08-202-186-20
9	18.2	58.7	1075	3	US-08-973-068-2
10	18.2	58.7	1089	3	US-08-202-186-23
11	18.2	58.7	1089	3	US-08-973-068-8
12	18.2	58.7	1110	1	US-08-202-186-14
13	17.8	57.4	2974	1	US-08-208-486-2
14	17.2	55.5	2284	4	US-08-193-792-1
15	17.2	55.5	4086	1	US-08-313-181-1
16	16.6	53.5	69	3	US-08-973-068-10
17	16.6	53.5	69	3	US-08-973-068-13
18	16.6	53.5	69	3	US-08-973-068-14
19	16.6	53.5	250	3	US-08-973-068-22
20	16.6	53.5	532	3	US-08-973-068-26
21	16.6	53.5	689	3	US-08-973-068-25
22	16.6	53.5	982	3	US-08-973-068-28
23	16.6	53.5	1018	1	US-08-202-186-22
24	16.6	53.5	1018	3	US-08-973-068-6
25	16.6	53.5	1043	1	US-08-202-186-21
26	16.6	53.5	1043	3	US-08-973-068-4
27	16.6	53.5	1103	1	US-08-202-186-16

c 28	16.6	53.5	1104	1	US-08-202-186-17	Sequence 17, Appl
c 29	16.6	53.5	1105	1	US-08-202-186-18	Sequence 18, Appl
c 30	16.6	53.5	1109	1	US-08-202-186-13	Sequence 13, Appl
c 31	16.6	53.5	1110	1	US-08-202-186-11	Sequence 11, Appl
c 32	16.6	53.5	1111	1	US-08-202-186-9	Sequence 9, Appl
c 33	16.6	53.5	1111	1	US-08-202-186-10	Sequence 10, Appl
c 34	16.6	53.5	1111	1	US-08-202-186-12	Sequence 12, Appl
c 35	16.6	53.5	1111	1	US-08-202-186-15	Sequence 15, Appl
c 36	16.6	53.5	1111	1	US-08-202-186-24	Sequence 24, Appl
c 37	16.6	53.5	1381	2	US-08-950-449A-13	Sequence 13, Appl
c 38	16.6	53.5	1423	1	US-08-469-421-13	Sequence 13, Appl
c 39	16.6	53.5	1423	1	US-08-250-975-13	Sequence 13, Appl
c 40	16.6	53.5	1423	2	US-08-605-002A-13	Sequence 13, Appl
c 41	16.6	53.5	1423	5	PCT-US94-10529-13	Sequence 13, Appl
c 42	16.6	53.5	3774	2	US-08-950-449A-11	Sequence 11, Appl
c 43	16.6	53.5	3813	1	US-08-469-421-11	Sequence 11, Appl
c 44	16.6	53.5	3813	1	US-08-250-975-11	Sequence 11, Appl
c 45	16.6	53.5	3813	2	US-08-605-002A-11	Sequence 11, Appl

ALIGNMENTS

```

RESULT 1
US-08-752-760A-1
; Sequence 1, Application US/08752760A
; Patent No. 5877011
; GENERAL INFORMATION:
; APPLICANT: Armentano, Donna
; APPLICANT: Gregory, Richard J.
; APPLICANT: Smith, Alan E.
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Baker & Botts, L.L.P.
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752.760A
; FILING DATE: 20-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seide, Rochelle K
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: A31385
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-705-5000
; TELEFAX: 212-705-5020
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35081 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-752-760A-1

```

Query Match 60.6% Score 18.8; DB 2; Length 35081;
Best Local Similarity 76.7%; Pred. No. 17;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

us-09-462-955-2.rni

Thu Oct 25 13:08:42 2001

QY 1 agccgcgggggtaatactagccccgcggc 30
 DB 15096 ACCGCTGGGTATTACTAGCCCGCAGC 15125

RESULT 2
 US-08-973-068-15/c
 ; Sequence 15, Application US/08973068
 ; Patent No. 6127604
 ; GENERAL INFORMATION:
 ; APPLICANT: Dale, James Langham
 ; APPLICANT: Harding, Robert Maxwell
 ; APPLICANT: Dugdale, Benjamin
 ; APPLICANT: Beetham, Peter Ronald
 ; APPLICANT: Hafner, Gregory John
 ; APPLICANT: Becker, Douglas Kenneth
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
 ; FILE REFERENCE: 09657/002001
 ; CURRENT APPLICATION NUMBER: US/08/973,068
 ; CURRENT FILING DATE: 1998-03-12
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335
 ; EARLIER FILING DATE: 1996-05-31
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 15
 ; LENGTH: 60
 ; TYPE: DNA
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 US-08-973-068-15

Query Match 58.7%; Score 18.2; DB 3; Length 60;
 Best Local Similarity 74.2%; Pred. No. 19;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagccccgcggc 31
 DB 47 AGCAGGGGGTAAATAATAGTCCCGCGTGT 17

RESULT 3
 US-08-973-068-12/c
 ; Sequence 12, Application US/08973068
 ; Patent No. 6127604
 ; GENERAL INFORMATION:
 ; APPLICANT: Dale, James Langham
 ; APPLICANT: Harding, Robert Maxwell
 ; APPLICANT: Dugdale, Benjamin
 ; APPLICANT: Beetham, Peter Ronald
 ; APPLICANT: Hafner, Gregory John
 ; APPLICANT: Becker, Douglas Kenneth
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
 ; FILE REFERENCE: 09657/002001
 ; CURRENT APPLICATION NUMBER: US/08/973,068
 ; CURRENT FILING DATE: 1998-03-12
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335
 ; EARLIER FILING DATE: 1996-05-31
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 69
 ; TYPE: DNA
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 US-08-973-068-12

Query Match 58.7%; Score 18.2; DB 3; Length 69;
 Best Local Similarity 74.2%; Pred. No. 19;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagccccgcggc 31
 DB 56 AGCGTGGGGTAAATAATAGTCCCGCAGC 26

RESULT 4
 US-08-973-068-30/c
 ; Sequence 30, Application US/08973068
 ; Patent No. 6127604
 ; GENERAL INFORMATION:
 ; APPLICANT: Dale, James Langham
 ; APPLICANT: Harding, Robert Maxwell
 ; APPLICANT: Dugdale, Benjamin
 ; APPLICANT: Beetham, Peter Ronald
 ; APPLICANT: Hafner, Gregory John
 ; APPLICANT: Becker, Douglas Kenneth
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
 ; FILE REFERENCE: 09657/002001
 ; CURRENT APPLICATION NUMBER: US/08/973,068
 ; CURRENT FILING DATE: 1998-03-12
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335
 ; EARLIER FILING DATE: 1996-05-31
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 30
 ; LENGTH: 352
 ; TYPE: DNA
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 US-08-973-068-30

Query Match 58.7%; Score 18.2; DB 3; Length 352;
 Best Local Similarity 74.2%; Pred. No. 22;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagccccgcggc 31
 DB 104 AGCAGGGGGTAAATAATAGTCCCGCGTGT 74

RESULT 5
 US-08-973-068-24/c
 ; Sequence 24, Application US/08973068
 ; Patent No. 6127604
 ; GENERAL INFORMATION:
 ; APPLICANT: Dale, James Langham
 ; APPLICANT: Harding, Robert Maxwell
 ; APPLICANT: Dugdale, Benjamin
 ; APPLICANT: Beetham, Peter Ronald
 ; APPLICANT: Hafner, Gregory John
 ; APPLICANT: Becker, Douglas Kenneth
 ; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
 ; FILE REFERENCE: 09657/002001
 ; CURRENT APPLICATION NUMBER: US/08/973,068
 ; CURRENT FILING DATE: 1998-03-12
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00335
 ; EARLIER FILING DATE: 1996-05-31
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 24
 ; LENGTH: 547
 ; TYPE: DNA
 ; ORGANISM: Banana Bunchy Top Virus (BBTV)
 US-08-973-068-24

Query Match 58.7%; Score 18.2; DB 3; Length 547;
 Best Local Similarity 74.2%; Pred. No. 22;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcgggggtaatactagccccgcggc 31
 DB 366 AGCGTGGGGTAAATAATAGTCCCGCAGC 336

RESULT 6

```

US-08-973-068-29/c
; Sequence 29, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; CURRENT APPLICATION NUMBER: US/08/973,068
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-29

Query Match          58.7%; Score 18.2; DB 3; Length 622;
Best Local Similarity 74.2%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgct 31
    ||| ||||| ||||| ||||| ||||| |||||
Db 374 AGCACGGGGGTAATAATAGTCCCCGTGCT 344

RESULT 7
US-08-973-068-27/c
; Sequence 27, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; CURRENT APPLICATION NUMBER: US/08/973,068
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
US-08-973-068-27

Query Match          58.7%; Score 18.2; DB 3; Length 624;
Best Local Similarity 74.2%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgct 31
    ||| ||||| ||||| ||||| ||||| |||||
Db 375 AGCACGGGGGTAATAATAGTCCCCGTGCT 345

RESULT 8
US-08-202-186-20/c
; Sequence 20, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1075 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-20

Query Match          58.7%; Score 18.2; DB 1; Length 1075;
Best Local Similarity 74.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggtaatactagcccccgcgct 31
    ||| ||||| ||||| ||||| ||||| |||||
Db 31 AGCGTGGGGTAATAATAGTCCCCAGCGCT 1

RESULT 9
US-08-973-068-2/c
; Sequence 2, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Beetham, Peter Ronald
; APPLICANT: Hafner, Gregory John
; APPLICANT: Becker, Douglas Kenneth
; TITLE OF INVENTION: INTERGENIC REGIONS OF BANANA BUNCHY TOP VIRUS
; FILE REFERENCE: 09657/002001
; CURRENT APPLICATION NUMBER: US/08/973,068
; CURRENT FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: PCT/AU96/00335
; EARLIER FILING DATE: 1996-05-31
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1075
; TYPE: DNA
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; FEATURE:
; NAME/KEY: CDS

```

RESULT 11
US-08-973-068-8/c
; Sequence 8, Application US/08973068
; Patent No. 6127604
; GENERAL INFORMATION:
; APPLICANT: Dale, James Langham
; APPLICANT: Harding, Robert Maxwell
; APPLICANT: Dugdale, Benjamin
; APPLICANT: Dugdale, Benjamin

Query Match 58.7%; Score 18.2; DB 1; Length 1110;
 Best Local Similarity 74.2%; Pred. No. 23;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggaataactagccccgcgcgcct 31
 ||| ||||| ||||| ||| ||||| |||
 Db 31 AGCGTGGGGTAATAATAAGTCCCGAGCGCT 1

RESULT 13
 US-08-208-486-2/c
 ; Sequence 2, Application US/08208486
 ; Patent No. 5389531
 ; GENERAL INFORMATION:
 ; APPLICANT: Ito, Junetsu
 ; APPLICANT: Yoo, Seung-Ku
 ; TITLE OF INVENTION: METHODS TO REPLICATE DNA IN VITRO USING
 ; TITLE OF INVENTION: PRD1-CATALYZED DNA REPLICATION SYSTEMS
 ; NUMBER OF SEQUENCES: 89
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cahill, Sutton & Thomas
 ; STREET: 155 Park One, 2141 E. Highland Ave.
 ; CITY: Phoenix
 ; STATE: Arizona
 ; COUNTRY: U.S.A.
 ; ZIP: 85016

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 5.25 inch, 1.2 Mb
 COMPUTER: Packard Bell (IBM PC/AT compatible)
 OPERATING SYSTEM: MS-Dos, Version 5.0
 SOFTWARE: WordPerfect Version 5.1

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/208,486
 FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/869,916

FILING DATE: April 14, 1992

APPLICATION NUMBER: Japan 240525/91

FILING DATE: August 26, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Janelle Faunce Raupp

REGISTRATION NUMBER: 30,485

REFERENCE/DOCKET NUMBER: #3954-A-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (602) 956-7000

TELEFAX: (602) 495-9475

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2974 base pairs

TYPE: nucleic acid

STRANDEDNESS: Double

TOPOLOGY: Linear

MOLECULE TYPE: Genomic DNA

FEATURE: 1-110 E CDS (L-TR)

FEATURE: 233-1012 E CDS (PRD1 terminal protein)

FEATURE: 1016-2677 E CDS (PRD1 DNA polymerase)

US-08-208-486-2

Query Match 57.4%; Score 17.8; DB 1; Length 2974;
 Best Local Similarity 75.9%; Pred. No. 37;
 Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 gccgcgggggaataactagccccgcgcgc 30
 ||||| ||||| ||||| ||| ||||| |||
 Db 2244 GCCGCGCTGTACTTCCGCCACTGC 2216

RESULT 14
 US-09-193-792-1/c
 ; Sequence 1, Application US/09193792B
 ; Patent No. 6180344

GENERAL INFORMATION:
 APPLICANT: Chen, Bin
 TITLE OF INVENTION: 5(Upstream Region Sequences of the MYOD1 Gene
 FILE REFERENCE: D6015
 CURRENT APPLICATION NUMBER: US/09/193,792B
 CURRENT FILING DATE: 1998-11-17
 PRIOR APPLICATION NUMBER: US 60/065,113
 PRIOR FILING DATE: 1997-11-18
 NUMBER OF SEQ ID NOS: 20
 SEQ ID NO 1

LENGTH: 2284

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: promoter

LOCATION: -1537..747

OTHER INFORMATION: 5(upstream promoter region of the human MYOD1 gene

US-09-193-792-1

Query Match 55.5%; Score 17.2; DB 4; Length 2284;
 Best Local Similarity 73.3%; Pred. No. 64;
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 agccgcggggggaataactagccccgcgcgc 30
 ||| ||||| ||||| ||| ||||| |||
 Db 2243 AGTCGCGCTGTAGTCTCGCGCGCGGC 2214

RESULT 15

US-08-313-181-1/c

Sequence 1, Application US/08313181

Patent No. 5681735

GENERAL INFORMATION:

APPLICANT: Emerson, Charles P.

APPLICANT: Goldhamer, David J.

TITLE OF INVENTION: Transcription Control Element for

TITLE OF INVENTION: Increasing Gene Expression in Myoblasts

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman

STREET: 1601 Market Street, Suite 720

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/313,181

FILING DATE: 07-OCT-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4086 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: not relevant

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:26 ; Search time 393.52 Seconds
(without alignments)
14.360 Million cell updates/sec

Title: US-09-462-955-3

Perfect score: 9

Sequence: 1 taatattac 9

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq:0601:*

- 1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	100.0	10	22	AAAF1274
2	9	100.0	17	20	AAV73850
3	9	100.0	18	14	AAQ48467
4	9	100.0	20	13	AAQ22027
5	9	100.0	20	13	AAQ22030
6	9	100.0	20	13	AAQ22035
7	9	100.0	20	14	AAQ41484
8	9	100.0	20	21	AAC65725
9	9	100.0	21	13	AAQ22039
10	9	100.0	21	20	AAQ06650
11	9	100.0	21	20	AAQ06641

c 12	9	100.0	21	20	AAQ01807	R. thuringiensis g
c 13	9	100.0	24	20	AAQ82184	Human hyd (h-Hyd)
c 14	9	100.0	24	20	AAQ82184	Human hyd (h-Hyd)
c 15	9	100.0	25	22	AAQ60809	S. cerevisiae MET1
c 16	9	100.0	26	21	AAQ08601	PCR primer #4 to a
c 17	9	100.0	26	21	AAQ08602	PCR primer #5 to a
c 18	9	100.0	26	21	AAQ47072	Primer SPC to mut
c 19	9	100.0	26	22	AAQ31180	Oligonucleotide Ss
c 20	9	100.0	27	21	AAQ40331	Human apolipoprote
c 21	9	100.0	34	21	AAQ43142	Primer for C.
c 22	9	100.0	35	18	AAQ94383	Primer sigK-4 to p
c 23	9	100.0	37	18	AAQ95210	Secondary alcohol
c 24	9	100.0	40	17	AAQ70616	Ligand L13 for per
c 25	9	100.0	42	10	AAQ97079	Sequence of C. tra
c 26	9	100.0	49	19	AAQ63501	Template YC-80 for
c 27	9	100.0	50	18	AAQ76820	Staphylococcus aur
c 28	9	100.0	55	20	AAQ15959	PCR primer Al-3 us
c 29	9	100.0	57	21	AAQ66232	Oligonucleotide us
c 30	9	100.0	60	20	AAQ83649	PCR primer SPH2 us
c 31	9	100.0	72	20	AAQ85096	Peripheral blood m
c 32	9	100.0	94	21	AAQ11499	Human secreted pro
c 33	9	100.0	95	21	AAQ28708	Human secreted pro
c 34	9	100.0	97	20	AAQ11476	Human secreted pro
c 35	9	100.0	118	21	AAQ28585	Sequence 300-7 ide
c 36	9	100.0	125	21	AAQ08016	Human secreted pro
c 37	9	100.0	126	21	AAQ45603	Human secreted exp
c 38	9	100.0	127	21	AAQ25885	Human secreted exp
c 39	9	100.0	129	16	AAQ21505	Human secreted pro
c 40	9	100.0	129	16	AAQ21505	Human gene signatu
c 41	9	100.0	132	21	AAQ29121	Human gene signatu
c 42	9	100.0	132	21	AAQ43018	Human secreted pro
c 43	9	100.0	134	16	AAQ21752	Human secreted exp
c 44	9	100.0	135	18	AAQ77559	Human gene signatu
c 45	9	100.0	135	21	AAQ22392	Staphylococcus aur
						Human secreted pro

ALIGNMENTS

RESULT 1

AAAF1274
ID AAF41274 standard; DNA; 10 BP.

XX AAF41274;

XX 23-MAR-2001 (first entry)

XX Yeast NORF gene SAGE tag oligonucleotide SEQ ID NO:8013.

XX Yeast; Saccharomyces cerevisiae; characterisation; cell cycle; NORF;
KW nor previously assigned open reading frame; nonannotated ORF; SAGE;
KW serial analysis of gene expression; antifungal; tag; identification;
KW linker; PCR primer; ds.

XX Saccharomyces cerevisiae.

XX WO200077214-A2.

XX 21-DEC-2000.

XX 14-JUN-2000; 2000WO-US16223.

XX 16-JUN-1999; 99US-0335032.

XX (UYJO) UNIV JOHNS HOPKINS.

XX Velculescu V, Vogelstein B, Kinzler K;

XX WPI; 2001-061874/07.

XX Yeast gene coding sequences comprising NORF genes with serial analysis
PT of gene expression (SAGE) tags, useful for studying, monitoring and
PT affecting phases of the cell cycle -

Thu Oct 25 13:08:45 2001

us-09-462-955-3.rng

XX Example; Page 286; 419pp.; English.

PS The present invention describes an isolated DNA molecule comprising a

XX coding sequence of a yeast gene selected from a group of 745 NORF (not

CC previously assigned open reading frame; or nonannotated ORF) genes

CC comprising a SAGE (serial analysis of gene expression) tag. Also

CC described are: (1) a method (M1) of using NORF genes to affect the cell

CC cycle comprising administering a NORF gene whose expression varies by at

CC least 10% between any two phases of the cell cycle selected from log

CC phase, S phase and G2/M; (2) a method (M2) for screening candidate

CC antifungal drugs comprising: (a) contacting a test substance with a

CC yeast cell; and (b) monitoring expression of a NORF gene whose

CC expression varies as in M1, where a test substance which modifies the

CC expression of the yeast gene is a candidate antifungal drug; (3) a method

CC (M3) for identifying human genes which are involved in cell cycle

CC progression comprising contacting human DNA with a probe which comprises

CC at least 10 contiguous nucleotides of a NORF gene whose expression varies

CC as in M1; and (4) a method (M4) for identifying a candidate drug as a

CC member of a class of drugs having a characteristic effect on gene

CC expression in a yeast cell comprising contacting a yeast cell with a

CC candidate drug and monitoring expression in the yeast cell of at least 1

CC NORF gene whose expression is affected by the class of drugs. The NORF

CC genes may be used to study, monitor and affect phases of the cell cycle,

CC the differentially expressed genes may be used as markers of phases of

CC the cell cycle. The methods may be used to identify candidate drugs which

CC affect the cell cycle and for identification of antifungal drugs.

CC AAF33268 to AAF44064 represent SAGE tags used in the exemplification of

CC the present invention. AAF33262 to AAF33267 represent linkers and PCR

CC primers used in the SAGE method, in the exemplification of the present

CC invention.

XX Sequence 10 BP; 4 A; 2 C; 0 G; 4 T; 0 other;

SQ

Query Match 100.0%; Score 9; DB 22; Length 10;

Best Local Similarity 100.0%; Pred. No. 9.5e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9

DB 2 taatattac 10

RESULT 2

AAV73850

ID AAV73850 standard; DNA; 17 BP.

XX

AC AAV73850;

XX

DT 25-FEB-1999 (first entry)

XX

DE C. albicans DNA target region for antisense inhibition.

XX

XX Antisense; inhibition; chiral phosphate linkage; reporter gene; drug;

KW RNase H activity; nuclease resistance; hybridisation; diagnostic;

KW cellular absorption; transport; enzymatic interaction; ss.

XX Candida albicans.

OS

XX US5852188-A.

PN

XX 22-DEC-1998.

PD

XX 19-APR-1996; 96US-0635009.

PF

XX 19-APR-1996; 96US-0635009.

PR

XX 11-JAN-1990; 90US-0463358.

PR

XX 13-AUG-1990; 90US-0566977.

PR

XX 11-JAN-1991; 91WO-US00243.

XX

PA (ISIS-) ISIS PHARM INC.

XX

PI Cook PD;

XX WPI; 1999-080505/07.

DR

XX New oligo-nucleotide(s) for e.g. testing anti-sense activity

PT comprise non-naturally occurring nucleoside unit and chiral

CC phosphate linkages

XX Disclosure; Column 12; 18pp; English.

PS

XX This sequence is used as a target sequence for a novel method to test

CC for antisense activity using an oligonucleotide comprising nucleoside

CC units linked via phosphate linkages in which at least one of the

CC nucleoside units is a non-naturally occurring nucleoside unit and at

CC least two of the nucleoside units are linked via chiral phosphate

CC linkages. The oligonucleotides can be used to test for antisense

CC activity using reporter genes in assays and to test antisense activity

CC against selected cellular target mRNA's in cultured cells. Some of the

CC oligonucleotides are useful for to elicit RNase H activity as a

CC termination event or to increase nuclease resistance. The

CC oligonucleotides are expected to exhibit one or more properties such as

CC hybridisation with target RNA's and DNA's, cellular absorption,

CC transport, or to improve enzymatic interaction without diminishing

CC existing properties giving improved, drugs, diagnostics and research

CC agents.

XX Sequence 17 BP; 6 A; 3 C; 2 G; 6 T; 0 other;

SQ

Query Match 100.0%; Score 9; DB 20; Length 17;

Best Local Similarity 100.0%; Pred. No. 9e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9

DB 7 taatattac 15

RESULT 3

AAQ48467/C

ID AAQ48467 standard; DNA; 18 BP.

XX

AC AAQ48467;

XX

DT 18-MAR-1994 (first entry)

XX

DE Nucleotides 970-987 of RepA TS mutant gene.

XX

XX Temperature sensitive replication; antibiotic resistance marker gene;

KW site-specific recombination; chromosomal integration; inactivation;

KW heterologous gene expression; thermosensitive plasmid;

KW suicide vector; Lactococcus; ss.

XX Synthetic.

OS

XX Key Location/Qualifiers

FT misc_difference 3

FT /*tag= a

FT /note= "wild-type G has been mutated to A

FT at position 972 of ORFA"

FT

FT misc_difference 8

FT /*tag= b

FT /note= "wild-type G has been mutated to A

FT at position 977 of ORFA"

FT

FT misc_difference 11

FT /*tag= c

FT /note= "wild-type G has been mutated to A

FT at position 980 of ORFA"

FT

FT misc_difference 18

FT /*tag= d

FT /note= "wild-type G has been mutated to A

FT at position 987 of ORFA"

XX

PN WO9318164-A.
 XX
 PD 16-SEP-1993.
 XX
 PF 12-MAR-1993; 93WO-FR00248.
 XX
 PR 13-MAR-1992; 92FR-0003034.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Gruss A, Maguin E;
 XX
 DR WPI; 1993-303478/38.
 DR P-PSDB; AAR41755.
 XX
 PT New bacterial plasmid contg. heat sensitive replication system -
 PT and marker gene, opt..capable of chromosomal integration, used to
 PT inactivate specific gene or introduce heterologous gene
 XX
 PS Claim 12; Fig 3; 73pp; French.
 XX
 CC Plasmid pWV01 is a cryptic plasmid originally isolated from
 CC Lactococcus lactis subsp. cremoris; it has a wide host range, being
 CC able to replicate in both gram-positive and gram-negative bacteria.
 CC There are 4 open reading frames in pWV01 with ORF-A coding for the
 CC replication protein RepA which cuts a DNA strand of ORI+. Plasmid
 CC pGK12 (Appl.Environ.Microbiol., 48; 726 (1984)) was derived from
 CC pWV01 and contains two antibiotic resistance marker genes.
 CC Mutagenesis of pGK12 with hydroxylamine resulted in isolation of a
 CC heat-stable mutant. Sequence analysis localised the mutations to 4
 CC sites within the RepA coding sequence, i.e. at nucleotides 972,
 CC 977, 980 and 987 of ORFA (see AAQ48467 for mutant sequence). The
 CC invention covers plasmids which comprise the temperature sensitive
 CC ORI; these plasmids can be used to inactivate host genes or to
 CC introduce heterologous genes into a bacterial host.
 XX
 SQ Sequence 18 BP; 9 A; 1 C; 1 G; 7 T; 0 other;

Query Match 100.0%; Score 9; DB 14; Length 18;
 Best Local Similarity 100.0%; Pred. No. 9e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
 Db 14 TAATATTAC 6

RESULT 4
 AAQ22027
 ID AAQ22027 standard; DNA; 20 BP.
 XX
 AC AAQ22027;
 XX
 DT 13-JUL-1992 (first entry)
 XX
 DE C. albicans CS antisense oligonucleotide analogue (1287).
 XX
 KW Candida albicans; diagnosis; infection; therapy; CS;
 KW chitin synthetase; ss.
 XX
 OS Synthetic.
 XX
 PN WO9203455-A.
 XX
 PD 05-MAR-1992.
 XX
 PF 15-AUG-1991; 91WO-US05802.
 XX
 PR 16-AUG-1990; 90US-0568672.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Hoke GD, Ecker DJ;
 XX
 DR WPI; 1992-096818/12.
 XX
 PT New oligo-nucleotide analogues as anti-sense agents - for the
 PT treatment and diagnosis of Candida infections
 XX
 PS Example 3; Page 14; 32pp; English.
 XX
 CC A phosphodiester and phosphorothioate analog of this sequence is
 CC disclosed. The oligonucleotides represented in AAQ22030-41 are tested
 CC in germ tube assay for inhibition of C. albicans. Oligonucleotide
 CC 2216, the phosphorothioate oligonucleotide analog of 2214 (AAQ22030),
 CC hybridisable with mRNA encoding Candida chitin synthetase, showed a
 CC greater inhibition of Candida germ tube formation relative to the
 CC other cpds.
 XX
 SQ Sequence 20 BP; 6 A; 3 C; 3 G; 8 T; 0 other;

PI Hoke GD, Ecker DJ;
 XX
 DR WPI; 1992-096818/12.
 XX
 PT New oligo-nucleotide analogues as anti-sense agents - for the
 PT treatment and diagnosis of Candida infections
 XX
 PS Example 1; Page 11; 32pp; English.
 XX
 CC The target RNA for this antisense oligonucleotide is CS
 CC (initiation of translation). The sequence modulates the
 CC activity of Candida RNA or DNA and thus can be used for treatment,
 CC diagnosis and research of Candida infections. Such antisense
 CC therapies are highly specific and do not cause toxic side effects
 CC to the host.
 XX
 SQ Sequence 20 BP; 6 A; 3 C; 3 G; 8 T; 0 other;

Query Match 100.0%; Score 9; DB 13; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.9e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
 Db 10 taatattac 18

RESULT 5
 AAQ22030
 ID AAQ22030 standard; DNA; 20 BP.
 XX
 AC AAQ22030;
 XX
 DT 13-JUL-1992 (first entry)
 XX
 DE C. albicans CS AUG antisense oligonucleotide analogue (2214).
 XX
 KW Candida albicans; diagnosis; infection; therapy; CS;
 KW chitin synthetase; phosphodiester; phosphorothioate; ss.
 XX
 OS Synthetic.
 XX
 PN WO9203455-A.
 XX
 PD 05-MAR-1992.
 XX
 PF 15-AUG-1991; 91WO-US05802.
 XX
 PR 16-AUG-1990; 90US-0568672.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Hoke GD, Ecker DJ;
 XX
 DR WPI; 1992-096818/12.
 XX
 PT New oligo-nucleotide analogues as anti-sense agents - for the
 PT treatment and diagnosis of Candida infections
 XX
 PS Example 3; Page 14; 32pp; English.
 XX
 CC A phosphodiester and phosphorothioate analog of this sequence is
 CC disclosed. The oligonucleotides represented in AAQ22030-41 are tested
 CC in germ tube assay for inhibition of C. albicans. Oligonucleotide
 CC 2216, the phosphorothioate oligonucleotide analog of 2214 (AAQ22030),
 CC hybridisable with mRNA encoding Candida chitin synthetase, showed a
 CC greater inhibition of Candida germ tube formation relative to the
 CC other cpds.
 XX
 SQ Sequence 20 BP; 6 A; 3 C; 3 G; 8 T; 0 other;

Query Match 100.0%; Score 9; DB 13; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.9e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
 Db 10 taatattac 18

RESULT 5
 AAQ22030
 ID AAQ22030 standard; DNA; 20 BP.
 XX
 AC AAQ22030;
 XX
 DT 13-JUL-1992 (first entry)
 XX
 DE C. albicans CS AUG antisense oligonucleotide analogue (2214).
 XX
 KW Candida albicans; diagnosis; infection; therapy; CS;
 KW chitin synthetase; phosphodiester; phosphorothioate; ss.
 XX
 OS Synthetic.
 XX
 PN WO9203455-A.
 XX
 PD 05-MAR-1992.
 XX
 PF 15-AUG-1991; 91WO-US05802.
 XX
 PR 16-AUG-1990; 90US-0568672.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Hoke GD, Ecker DJ;
 XX
 DR WPI; 1992-096818/12.
 XX
 PT New oligo-nucleotide analogues as anti-sense agents - for the
 PT treatment and diagnosis of Candida infections
 XX
 PS Example 3; Page 14; 32pp; English.
 XX
 CC A phosphodiester and phosphorothioate analog of this sequence is
 CC disclosed. The oligonucleotides represented in AAQ22030-41 are tested
 CC in germ tube assay for inhibition of C. albicans. Oligonucleotide
 CC 2216, the phosphorothioate oligonucleotide analog of 2214 (AAQ22030),
 CC hybridisable with mRNA encoding Candida chitin synthetase, showed a
 CC greater inhibition of Candida germ tube formation relative to the
 CC other cpds.
 XX
 SQ Sequence 20 BP; 6 A; 3 C; 3 G; 8 T; 0 other;

XX

DE Target region for *Candida albicans*.
XX

DR WPI: 2000-686014/67.
 XX
 PT Antisense compound 8-30 nucleobases in length targeted to a start codon
 PT of the coding region of human p13 kinase p10beta, useful for
 PT inhibiting the expression of the human polynucleotide -
 XX
 PS Claim 14: Column 61-62: 34pp; English.
 XX
 CC This invention describes a novel antisense compound (I) 8-30 nucleobases
 CC in length targeted to a start codon or nucleobases 4-3174 of the coding
 CC region of human p13 kinase p10beta (II), in which (I) specifically
 CC hybridizes with and inhibits the expression of (II). The products of the
 CC invention have cytostatic, antiinflammatory and antiinfective activity.
 CC (I) is useful for inhibiting the expression of (II) in human cells or
 CC tissues. The antisense compound can be utilized for diagnostics,
 CC therapeutics, prophylaxis and as research reagents and kits. The
 CC antisense compounds may also be useful prophylactically, e.g. to prevent
 CC or delay infection, inflammation or tumor formation. The antisense
 CC compounds are useful for research and diagnostics, because these
 CC compounds hybridize to nucleic acids encoding (II).
 XX
 SQ Sequence 20 BP; 5 A; 3 C; 3 G; 9 T; 0 other;
 Query Match 100.0%; Score 9; DB 21; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.9e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 taatattac 9
 Db 10 TAAATATTAC 2
 RESULT 9
 ID AAQ22039 standard; DNA; 21 BP.
 AC AAQ22039;
 DT 13-JUL-1992 (first entry)
 XX
 DE C. albicans CS AUG antisense oligonucleotide analogue (3150).
 XX
 KW Candida albicans; diagnosis; infection; therapy; CS;
 KW chitin synthetase; phosphorothioate; ss.
 OS Synthetic.
 XX
 PN WO9203455-A.
 XX
 PD 05-MAR-1992.
 XX
 PF 15-AUG-1991; 91WO-US05802.
 XX
 PR 16-AUG-1990; 90US-0568672.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Hoke GD, Ecker DJ;
 XX
 WPI: 1992-096818/12.
 XX
 PT New oligo-nucleotide analogues as anti-sense agents - for the
 PT treatment and diagnosis of Candida infections
 XX
 PS Example 3; Page 14; 32pp; English.
 XX
 CC The sequence is a phosphorothioate analog.
 CC The oligonucleotides represented in AAQ22030-41 are tested
 CC in germ tube assay for inhibition of C. albicans. Oligonucleotide
 CC 2216, the phosphorothioate oligonucleotide analog of 2214 (AAQ22030),
 CC hybridisable with mRNA encoding Candida chitin synthetase, showed a
 CC greater inhibition of Candida germ tube formation relative to the

CC other cpds.
 XX
 SQ Sequence 21 BP; 8 A; 4 C; 2 G; 7 T; 0 other;
 Query Match 100.0%; Score 9; DB 13; Length 21;
 Best Local Similarity 100.0%; Pred. No. 8.8e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 taatattac 9
 Db 6 taatattac 14
 RESULT 10
 ID AAX06650/c
 XX AAX06650 standard; DNA; 21 BP.
 AC AAX06650;
 XX
 DT 08-APR-1999 (first entry)
 XX
 DE B. anthracis gyrB gene specific primer.
 XX
 KW Nucleic acid amplification; detection; Bacillus anthracis; PCR; gyrB;
 KW DNA gyrase subunit B; PCR primer; ss.
 XX
 OS Synthetic.
 OS Bacillus anthracis.
 XX
 PN JP11004693-A.
 XX
 PD 12-JAN-1999.
 XX
 PF 29-SEP-1997; 97JP-0264059.
 XX
 PR 21-APR-1997; 97JP-0102843.
 PR 24-MAR-1997; 97JP-0069900.
 XX
 PA (NIUS) NIPPON SUISAN KAISHA LTD.
 XX
 WPI: 1999-135786/12.
 XX
 PT New oligonucleotide specific for Bacillus anthracis - amplifies
 PT Bacillus anthracis DNA gyrase subunit B in Bacillus anthracis
 XX
 PS Claim 4; Page 10; 11pp; Japanese.
 XX
 CC The invention relates to an oligonucleotide primer for amplification of
 CC a nucleic-acid target sequence characteristic of Bacillus anthracis. The
 CC oligonucleotide acts as a primer for amplifying the DNA gyrase subunit B
 CC (gyrB) gene for the detection of B. anthracis. A gyrB gene fragment
 CC specific to Bacillus anthracis can be detected using PCR without
 CC extraction of DNA. The method is useful for specific detection of B.
 CC anthracis and distinguishing from B. thuringiensis and B. cereus
 CC strains in a sample. The present sequence represents a primer specific
 CC for the B. anthracis gyrB gene.
 XX
 SQ Sequence 21 BP; 9 A; 3 C; 3 G; 6 T; 0 other;
 Query Match 100.0%; Score 9; DB 20; Length 21;
 Best Local Similarity 100.0%; Pred. No. 8.8e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 taatattac 9
 Db 13 TAAATATTAC 5
 RESULT 11
 ID AAX06641/c
 XX AAX06641 standard; DNA; 21 BP.

Thu Oct 25 13:08:45 2001

us-09-462-955-3.rng

```

XX AAX06641;
AC
XX
XX 08-APR-1999 (first entry)
DT
XX
DE B. anthracis gyrB gene specific primer.
XX
XX Nucleic acid amplification; detection; Bacillus cereus; PCR; gyrB;
KW DNA gyrase subunit B; PCR primer; ss.
XX
XX Synthetic.
OS Bacillus anthracis.
XX
XX JP11004691-A.
PN
XX
XX 12-JAN-1999.
PD
XX
XX 29-SEP-1997; 97JP-0264057.
PF
XX 21-APR-1997; 97JP-0102843.
PR
XX 24-MAR-1997; 97JP-0069900.
PA (NIUS ) NIPPON SUISAN KAISHA LTD.
XX
XX WPI; 1999-135785/12.
DR
XX New oligonucleotide specific for Bacillus cereus - amplifies
PT Bacillus cereus DNA gyrase subunit B in Bacillus cereus
XX
XX Examples; Page 10; 11pp; Japanese.
PS
XX The invention relates to an oligonucleotide primer for amplification of
CC a nucleic-acid target sequence characteristic of Bacillus cereus. The
CC oligonucleotide acts as a primer for amplifying the DNA gyrase subunit B
CC (gyrB) gene for the detection of B. cereus. A gyrB gene fragment specific
CC to Bacillus cereus can be detected using PCR without extraction of DNA.
CC The method is useful for specific detection of B. cereus and
CC distinguishing it from B. thuringiensis and B. anthracis strains in
CC a sample. The present sequence represents a primer specific for the
CC B. anthracis gyrB gene.
XX
XX Sequence 21 BP; 9 A; 3 C; 3 G; 6 T; 0 other;
SQ
Query Match 100.0%; Score 9; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taatattac 9
DB 13 TAATATTAC 5

RESULT 12
AAX01807/c
ID AAX01807 standard; DNA; 21 BP.
XX
XX AAX01807;
XX
XX 09-APR-1999 (first entry)
DT
DE B. thuringiensis gyrB PCR primer #5.
XX
XX gyrB protein; detection; PCR primer; soil; food; ss.
XX
XX Synthetic.
OS Bacillus thuringiensis.
XX
XX JP11004692-A.
PN
XX 12-JAN-1999.
PD
XX 29-SEP-1997; 97JP-0264058.
PF

```

```

XX 21-APR-1997; 97JP-0102843.
PR 24-MAR-1997; 97JP-0069900.
XX
XX (NIUS ) NIPPON SUISAN KAISHA LTD.
XX
XX WPI; 1999-145652/13.
DR
XX Oligonucleotide for Bacillus thuringiensis detection - has nucleic
PT acid sequence obtained from a sequence characteristic for the
PT microorganism
XX
XX Example 2; Page 10; 11pp; Japanese.
PS
XX AAX01803-X01808 are PCR primers used to amplify the Bacillus
CC thuringiensis gyrB genes represented in AAX01800-X01802. These primers
CC can be used in methods to detect the presence of B. thuringiensis in soil
CC or foodstuffs. The primers used in this method can distinguishing
CC between B. thuringiensis, B. cereus and B. anthracis.
XX
XX Sequence 21 BP; 9 A; 3 C; 3 G; 6 T; 0 other;
SQ
Query Match 100.0%; Score 9; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.8e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taatattac 9
DB 13 TAATATTAC 5

RESULT 13
AAX82184
ID AAX82184 standard; DNA; 24 BP.
XX
XX AAX82184;
XX
XX 17-AUG-1999 (first entry)
DT
DE Human hyd (h-Hyd) protein DNA amplifying primer.
XX
XX Human; hyd protein; h-Hyd; hect domain; E6AP; cell growth; heart;
KW intracellular protein decomposition; sperm cell; PCR primer; ss.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX JP11146786-A.
PN
XX 02-JUN-1999.
PD
XX 17-NOV-1997; 97JP-0314935.
PF
XX 17-NOV-1997; 97JP-0314935.
PR
XX (SOME ) SUMITOMO ELECTRIC IND CO.
XX
XX WPI; 1999-378996/32.
XX
XX New h-Hyd protein - and DNA encoding it
PT
XX Examples; Page 5; 20pp; Japanese.
PS
XX The invention relates to a human-hyd (h-Hyd) protein having hect domains
CC and homology to the C-terminal domain of E6AP. h-Hyd is presumed to
CC participate in intracellular protein decomposition and cell growth
CC especially in heart and sperm cells. Sequences AAX82184-187 represent
CC PCR primers for amplifying the h-Hyd protein DNA.
XX
XX Sequence 24 BP; 11 A; 5 C; 2 G; 6 T; 0 other;
SQ

```

```

Query Match      100.0%; Score 9; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 4 taatattac 12

RESULT 14
ID AAX82184/c
XX AAX82184 standard; DNA; 24 BP.
XX AC AAX82184;
XX DT 17-AUG-1999 (first entry)
XX DE Human hyd (h-Hyd) protein DNA amplifying primer.
XX KW Human; hyd protein; h-Hyd; hect domain; E6AP; cell growth; heart;
XX KW intracellular protein decomposition; sperm cell; PCR primer; ss.
XX OS Synthetic.
XX OS Homo sapiens.
XX PN JP11146786-A.
XX PD 02-JUN-1999.
XX PF 17-NOV-1997; 97JP-0314935.
XX PR 17-NOV-1997; 97JP-0314935.
XX PA (SUME ) SUMITOMO ELECTRIC IND CO.
XX DR WPI; 1999-378996/32.
XX PT New h-Hyd protein - and DNA encoding it
XX PS Examples; Page 5; 20pp; Japanese.
XX CC The invention relates to a human-hyd (h-Hyd) protein having hect domains
XX CC and homology to the C-terminal domain of E6AP; h-Hyd is presumed to
XX CC participate in intracellular protein decomposition and cell growth
XX CC especially in heart and sperm cells. Sequences AAX82184-187 represent
XX CC PCR primers for amplifying the h-Hyd protein DNA.
XX SQ Sequence 24 BP; 11 A; 5 C; 2 G; 6 T; 0 other;

Query Match      100.0%; Score 9; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 11 TAATATTAC 3

RESULT 15
ID AAF60809
XX AAF60809 standard; DNA; 25 BP.
XX AC AAF60809;
XX DT 04-MAY-2001 (first entry)
XX DE S. cerevisiae MET14 PCR primer SEQ ID 11.
XX KW Microorganism; sulfite production; sulfite cycle; food production; wine;
XX KW beer; taste stabilizer; non-volatile complex; carbonyl compound; yeast;
XX KW oxidation; fermentation; PCR primer; ss.

```

```

OS Saccharomyces cerevisiae.
XX DE19923950-A1.
XX PN 25-JAN-2001.
XX PD 25-MAY-1999; 99DE-1023950.
XX PF 25-MAY-1999; 99DE-1023950.
XX PR (STAHL/) STAHL U.
XX PA Donalis U, Stahl U;
XX PI WPI; 2001-148153/16.
XX DR
XX PT New microorganisms that produce high sulfite levels at a late stage in
XX PT their growth, useful for producing beer, prevent development of
XX PT off-flavors by oxidation
XX PS Example 4; Page 8; 14pp; German.
XX CC This invention describes novel microorganisms (A) able to produce delayed
XX CC and large amounts of sulfite. The microorganisms comprise a DNA construct
XX CC (I) containing one or more genes (II) involved in the sulfite cycle under
XX CC the control of a promoter. The high sulfite concentration appears at a
XX CC late stage of substrate utilization, in the stationary growth phase, in
XX CC the last third of the exponential growth phase or at a cell density of
XX CC 60-90% of that achieved in the growth phase. (A), particularly bacteria
XX CC and/or yeast, are used for the production of foods, wine, beer or desired
XX CC metabolic end products. Sulfite stabilizes the taste of beer by forming
XX CC non-volatile complexes with carbonyl compounds (formed by oxidation and
XX CC responsible for off-flavors) and by preventing oxidation (reducing
XX CC agent). (A) produce significant amounts of sulfite only at a late stage
XX CC in its growth, after the fermentation product has formed, avoiding
XX CC premature formation of complexes and eliminating the need to add sulfite
XX CC to the finished product.
XX SQ Sequence 25 BP; 7 A; 7 C; 3 G; 8 T; 0 other;

Query Match      100.0%; Score 9; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 14 taatattac 22

```

Search completed: October 24, 2001, 10:07:27
Job time: 632 sec

us-09-462-955-3.rng

Thu Oct 25 13:08:45 2001

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:52 ; Search time 5479.82 Seconds
(without alignments)
15.525 Million cell updates/sec

Title: US-09-462-955-3
Perfect score: 9
Sequence: 1 taatattac 9

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est33:*
- 26: gb_est34:*
- 27: gb_est35:*
- 28: gb_est36:*
- 29: gb_est37:*
- 30: gb_est38:*
- 31: gb_est39:*
- 32: gb_est40:*
- 33: em_estba:*
- 34: em_estfun:*
- 35: em_esthum1:*
- 36: em_esthum2:*
- 37: em_esthum3:*
- 38: em_esthum4:*
- 39: em_esthum5:*
- 40: em_esthum6:*
- 41: em_esthum7:*
- 42: em_esthum8:*
- 43: em_esthum9:*

- 44: em_esthum10:*
- 45: em_esthum11:*
- 46: em_esthum12:*
- 47: em_esthum13:*
- 48: em_esthum14:*
- 49: em_esthum15:*
- 50: em_esthum16:*
- 51: em_esthum17:*
- 52: em_esthum18:*
- 53: em_esthum19:*
- 54: em_esthum20:*
- 55: em_esthum21:*
- 56: em_esthum22:*
- 57: em_esthum23:*
- 58: em_esthum24:*
- 59: em_esthum25:*
- 60: em_esthum26:*
- 61: em_esthum27:*
- 62: em_esthum28:*
- 63: em_estin1:*
- 64: em_estin2:*
- 65: em_estin3:*
- 66: em_estin4:*
- 67: em_estin5:*
- 68: em_estomi:*
- 69: em_estom2:*
- 70: em_estov1:*
- 71: em_estov2:*
- 72: em_estpl1:*
- 73: em_estpl2:*
- 74: em_estpl3:*
- 75: em_estpl4:*
- 76: em_estpl5:*
- 77: em_estpl6:*
- 78: em_estpl7:*
- 79: em_estpl8:*
- 80: em_estpl9:*
- 81: em_estpl10:*
- 82: em_estrol:*
- 83: em_estro2:*
- 84: em_estro3:*
- 85: em_estro4:*
- 86: em_estro5:*
- 87: em_estro6:*
- 88: em_estro7:*
- 89: em_estro8:*
- 90: em_estro9:*
- 91: em_estro10:*
- 92: em_estro11:*
- 93: em_estro12:*
- 94: em_estro13:*
- 95: em_estro14:*
- 96: em_estro15:*
- 97: em_estro16:*
- 98: em_estro17:*
- 99: em_estro18:*
- 100: em_estro19:*
- 101: em_estro20:*
- 102: gb_est25:*
- 103: gb_est26:*
- 104: gb_est27:*
- 105: gb_est28:*
- 106: gb_est29:*
- 107: gb_est30:*
- 108: gb_est31:*
- 109: gb_est32:*
- 110: gb_est41:*
- 111: gb_est42:*
- 112: gb_est43:*
- 113: gb_est44:*
- 114: gb_est45:*
- 115: gb_est46:*
- 116: gb_est47:*

190:	gb_estll10:*
191:	gb_estll1:*
192:	gb_hcc:*
193:	em_gss_fun1:*
194:	em_gss_hum1:*
195:	em_gss_hum2:*
196:	em_gss_hum3:*
197:	em_gss_hum4:*
198:	em_gss_hum5:*
199:	em_gss_hum6:*
200:	em_gss_hum7:*
201:	em_gss_hum8:*
202:	em_gss_hum9:*
203:	em_gss_inv1:*
204:	em_gss_inv2:*
205:	em_gss_inv3:*
206:	em_gss_other:*
207:	em_gss_pln1:*
208:	em_gss_pln2:*
209:	em_gss_pro:*
210:	em_gss_rpd1:*
211:	em_gss_rpd2:*
212:	em_gss_rpd3:*
213:	em_gss_rpd4:*
214:	em_gss_rpd5:*
215:	em_gss_vrt1:*
216:	em_gss_vrt2:*
217:	em_gss_vrt3:*
218:	gb_gss1:*
219:	gb_gss2:*
220:	gb_gss3:*
221:	gb_gss4:*
222:	gb_gss5:*
223:	gb_gss6:*
224:	gb_gss7:*
225:	gb_gss8:*
226:	gb_gss9:*
227:	gb_gss10:*
228:	gb_gss11:*
229:	gb_gss12:*
230:	gb_gss13:*
231:	gb_gss14:*
232:	gb_gss15:*
233:	gb_gss16:*
234:	gb_gss17:*
235:	gb_gss18:*
236:	gb_gss19:*
237:	gb_gss20:*
238:	gb_gss21:*
239:	gb_gss22:*
240:	gb_gss23:*
241:	gb_gss24:*
242:	gb_gss25:*
243:	gb_gss26:*
244:	gb_gss27:*
245:	gb_gss28:*
246:	gb_gss29:*
247:	gb_gss30:*
248:	gb_gss31:*
249:	gb_gss32:*
250:	gb_gss33:*
251:	gb_gss34:*
252:	em_gss_inv4:*
253:	em_gss_rpd6:*
254:	em_gss_rpd7:*
255:	em_gss_rpd8:*
256:	gb_gss35:*
257:	gb_gss36:*
258:	gb_gss37:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	9	100.0	24	246	AZ595661 1M0408K24
2	25	249	AZ803817	25	AZ803817 2M0064H22
3	4	9	100.0	28	AZ595666 1M0408L23
4	9	100.0	31	246	AZ595666 1M0408L23
5	9	100.0	31	246	AZ485621 1M0313H08
6	9	100.0	34	24	AI784676 ta03f03.x
7	9	100.0	34	246	AZ589161 1M0397O22
8	9	100.0	36	244	AZ474278 1M0290F15
9	9	100.0	51	106	AU007795 AU007795
10	9	100.0	58	4	AA284056 2S60D05.s
11	9	100.0	58	121	AW827116 xn13e07.y
12	9	100.0	64	244	AZ471237 1M0285C13
13	9	100.0	66	246	AZ587595 1M0395E24
14	9	100.0	68	164	BE191328 sn71b06.y
15	9	100.0	69	249	AZ755547 ex01c06.x
16	9	100.0	71	174	BG226871 kp99a04.y
17	9	100.0	72	256	AZ929450 479.dif26
18	9	100.0	73	102	AI833171 at65e03.x
19	9	100.0	73	102	AI833171 at65e03.x
20	9	100.0	73	118	AW600128 SWL4CAK10
21	9	100.0	73	145	BF145150 SWL4CAK10
22	9	100.0	81	156	C11376 SWOV13CAN
23	9	100.0	81	243	AZ427672 C11376 Yuj1
24	9	100.0	82	244	AZ489440 1M0209G15
25	9	100.0	84	162	BE022876 sm77f04.y
26	9	100.0	84	249	AZ783162 2M0024D08
27	9	100.0	85	4	AA273976 vb94a07.r
28	9	100.0	85	7	AA445562 vf61n07.r
29	9	100.0	86	120	AW765103 ga88b07.y
30	9	100.0	87	13	AA865195 oh28c05.s
31	9	100.0	87	241	AZ318000 1M0036K24
32	9	100.0	88	4	AA276053 vc36h08.r
33	9	100.0	88	162	BE057653 sn05c09.y
34	9	100.0	88	244	AZ494555 1M0329H20
35	9	100.0	88	246	AZ602035 1M0420A07
36	9	100.0	89	13	AA869288 vq50a11.r
37	9	100.0	91	13	AA910700 ok89e10.s
38	9	100.0	94	11	AA708850 z158b03.s
39	9	100.0	94	12	AA841919 MB3D6V8H1
40	9	100.0	94	18	AI321285 d6e03nm.r
41	9	100.0	94	241	AZ322497 1M0043008
42	9	100.0	97	8	AA498371 vh43c01.r
43	9	100.0	97	8	AA498371 vh43c01.r
44	9	100.0	97	111	AW099103 sd34e03.y
45	9	100.0	97	122	AW909872 ur87d02.x

ALIGNMENTS

RESULT 1	AZ595661	24 bp	DNA	13-DEC-2000
LOCUS	1M0408K24F	Mouse 10kb plasmid	UUGCLM library	Mus musculus genomic
DEFINITION	clone UUGCLM0408K24 F, DNA sequence.			
ACCESSION	AZ595661.1	GI:11717851		
VERSION	GSS.			
KEYWORDS	house mouse.			
SOURCE	Mus musculus			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 24)			
REFERENCE	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.			

TITLE
JOURNAL
COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert H. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0408 row: K column: 24
Seq primer: CGTTGTAACACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 24.

FEATURES
source

Location/Qualifiers
1..24
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0408K24"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (G14732114/gb1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
ORIGIN

6 a 3 c 0 g 15 t

Query Match 100.0%; Score 9; DB 246; Length 24;
Best Local Similarity 100.0%; Pred. No. 7.3e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||
Db 3 TAATATTAC 11

RESULT 2
AZ803817/C

LOCUS
DEFINITION
2M0064H22F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0064H22 F, DNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AZ803817 25 bp DNA GSS 16-FEB-2001
AZ803817
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 25)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0408 row: L column: 23
 Seq primer: CGTTGTAACGACGCGCCACT
 Class: plasmid ends
 High quality sequence stop: 28.
 Location/Qualifiers
 1. .28
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0408L23"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

FEATURES
 source
 1. .28
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0408L23"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match 100.0%; Score 9; DB 246; Length 28;
 Best Local Similarity 100.0%; Pred. No. 7.3e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
 Db 28 TAATATTAC 20

RESULT 4
 AZ485621 31 bp DNA GSS 05-OCT-2000
 LOCUS
 DEFINITION
 1M0313H08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0313H08 F, DNA sequence.
 AZ485621
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 31)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0064 row: H column: 22
 Seq primer: CGTTGTAACGACGCGCCACT
 Class: plasmid ends
 High quality sequence stop: 25.
 Location/Qualifiers
 1. .25
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0064H22"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

FEATURES
 source
 1. .25
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0064H22"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

Query Match 100.0%; Score 9; DB 249; Length 25;
 Best Local Similarity 100.0%; Pred. No. 7.3e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatattac 9
 Db 9 TAATATTAC 1

RESULT 3
 AZ595666 28 bp DNA GSS 13-DEC-2000
 LOCUS
 DEFINITION
 1M0408L23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0408L23 F, DNA sequence.
 AZ595666
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 28)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

Thu Oct 25 13:08:48 2001

Plate: 0290 row: F column: 15
 Seq primer: CACACAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 36.
 Location/Qualifiers

FEATURES

source

1. .36
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0290F15"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 14 a 7 c 6 g 9 t
 ORIGIN

Query Match 100.0%; Score 9; DB 244; Length 36;
 Best Local Similarity 100.0%; Pred. No. 7.2e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taattattac 9
 |||||
 Db 11 TAATATTAC 19

RESULT 8
 AZ474278 51 bp mRNA EST 31-JUL-1998
 LOCUS
 DEFINITION Schizosaccharomyces pombe late log phase cDNA
 Schizosaccharomyces pombe cDNA clone spc02513, mRNA sequence.
 ACCESSION AU007795
 VERSION AU007795.1 GI:3344253
 KEYWORDS EST.
 SOURCE fission yeast.
 ORGANISM Schizosaccharomyces pombe
 Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 Schizosaccharomycetales; Schizosaccharomycetaceae;
 Schizosaccharomycetes.
 REFERENCE 1 (bases 1 to 51)
 AUTHORS Morimyo, M. and Mita, K.
 TITLE Identification of expressed sequence tags of Schizosaccharomycetes pombe
 JOURNAL Unpublished (1998)
 COMMENT Contact: Mitsuoki Morimyo
 Genome Research Group
 National Institute of Radiological Sciences
 9-1, Anagawa-4-chome, Inage-ku, Chiba 263-8555, Japan
 Email: morimyo@nirs.go.jp.
 Location/Qualifiers

FEATURES source
 1. 51
 /organism="Schizosaccharomyces pombe"
 /strain="972"
 /db_xref="taxon:4896"

Seq primer: CACACAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 34.
 Location/Qualifiers

FEATURES

source

1. .34
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0397022"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gll4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 13 a 8 c 4 g 9 t
 ORIGIN

Query Match 100.0%; Score 9; DB 246; Length 34;
 Best Local Similarity 100.0%; Pred. No. 7.2e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taattattac 9
 |||||
 Db 6 TAATATTAC 14

RESULT 7
 AZ474278 36 bp DNA GSS 04-OCT-2000
 LOCUS
 DEFINITION LM0290F15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0290F15 R, DNA sequence.
 ACCESSION AZ474278
 VERSION AZ474278.1 GI:10632403
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 36)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D. Weiss, R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00

```

/clone="spc02513"
/clone_lib="Schizosaccharomyces pombe late log phase cDNA"
/sex="h minus"
/note="Vector: M13mp19; The cDNA library of
Schizosaccharomyces pombe was prepared by cloning cDNA
into the SmaI site of M13mp19 DNA and the direction of DNA
sequences was not always from 5' to 3'. The cDNA data of
Schizosaccharomyces pombe are available for searching on
the World Wide Web. (URL, http://www.nirs.go.jp)"
BASE COUNT      22 a      3 c      7 g      19 t
ORIGIN

```

```

Query Match      100.0%; Score 9; DB 106; Length 51;
Best Local Similarity 100.0%; Pred. No. 7.1e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
    |||||
Db 17 TAATATTAC 9

```

```

RESULT 9
AA284056      58 bp      mRNA      EST      15-DEC-1997
LOCUS      zs60d05.s1 Stratagene schizo brain s11 Homo sapiens cDNA clone
DEFINITION      IMAGE:701865 3', mRNA sequence.
ACCESSION      AA284056
VERSION      AA284056.1 GI:1928337
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 58)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,
T., Waterston, R., and Wilson, R.
WashU-Merck EST Project 1997
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyT not found
Seq primer: -41m13 fwd. Et from Amersham.

```

```

FEATURES
    source
        1..58
            /organism="Homo sapiens"
            /db_xref="GDB:5742536"
            /db_xref="taxon:9606"
            /clone="IMAGE:701865"
            /clone_lib="Stratagene schizo brain s11"
            /sex="male"
            /tissue_type="schizophrenic brain s-11 frontal lobe"
            /dev_stage="34 years old"
            /lab_host="SOLR (kanamycin resistant)"
            /note="Vector: Bluescript SK-; Site1: EcoRI; Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4Kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric
Diseases (Unpublished) Stanley Neurovirology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT      22 a      5 c      16 g      15 t

```

ORIGIN

```

Query Match      100.0%; Score 9; DB 4; Length 58;
Best Local Similarity 100.0%; Pred. No. 7.1e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
    |||||
Db 8 TAATATTAC 16

```

```

RESULT 10
AW827116      58 bp      mRNA      EST      18-MAY-2000
LOCUS      xn13e07.y1 NCI_CGAP_Li5 Homo sapiens cDNA clone IMAGE:2693604 5',
DEFINITION      mRNA sequence.
ACCESSION      AW827116
VERSION      AW827116.1 GI:7920970
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 58)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Putative full length read
The vector to vector length is 59
Seq primer: -40RP from Gibco.

```

FEATURES

```

    source
        1..58
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2693604"
            /clone_lib="NCI_CGAP_Li5"
            /tissue_type="hepatic adenoma"
            /lab_host="DH10B"
            /note="Organ: liver; Vector: pCMV-SPORT4; Site1: Salt;
            Site2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Average insert size 0.8 kb."
            /dev_stage="7 g"
            /sex="male"
            /tissue_type="schizophrenic brain s-11 frontal lobe"
            /dev_stage="34 years old"
            /lab_host="SOLR (kanamycin resistant)"
            /note="Vector: Bluescript SK-; Site1: EcoRI; Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II Vector. Mass
excised. Custom library. Avg insert length 1.4Kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric
Diseases (Unpublished) Stanley Neurovirology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT      24 a      10 c      7 g      17 t
ORIGIN

```

```

Query Match      100.0%; Score 9; DB 121; Length 58;
Best Local Similarity 100.0%; Pred. No. 7.1e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
    |||||
Db 8 TAATATTAC 16

```

```

RESULT 11
AZ471237      64 bp      DNA      GSS      04-OCT-2000
LOCUS      1M0285C13R Mouse 10kb plasmid U06C1M library Mus musculus genomic
DEFINITION      clone U06C1M0285C13 R, DNA sequence.
ACCESSION      AZ471237
VERSION      AZ471237.1 GI:10629362

```

Thu Oct 25 13:08:48 2001

GSS.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 64)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0285 row: C column: 13
 Seq primer: CACACAGGACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 64.
 Location/Qualifiers
 1. .64
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0285C13"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 17 a 1 c 19 g 27 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 244; Length 64;
 Best Local Similarity 100.0%; Pred. No. 7e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 taattacc 9
 Db 50 TAATATTAC 42
 RESULT 12
 AZ587595 66 bp DNA GSS 13-DEC-2000
 LOCUS 1M0395E24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC1M0395E24 F, DNA sequence.
 ACCESSION AZ587595

AZ587595.1 GI:11709785
 GSS.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 66)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0395 row: E column: 24
 Seq primer: CGTTGTAACGACGCCAGT
 Class: plasmid ends
 High quality sequence stop: 66.
 Location/Qualifiers
 1. .66
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0395E24"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 34 a 11 c 3 g 18 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 246; Length 66;
 Best Local Similarity 100.0%; Pred. No. 7e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 taattacc 9
 Db 40 TAATATTAC 48
 RESULT 13
 BE191328/c 68 bp mRNA EST 22-JUN-2000
 LOCUS BE191328
 DEFINITION sn71b06.y1 Cm-cl038 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl038-12 5', mRNA sequence.

```

ACCESSION BE191328
VERSION BE191328.1 GI:8670221
KEYWORDS EST
SOURCE soybean.
ORGANISM Glycine max

Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 68)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
, Y., Person, B., Swaller, J., Gibbons, M., Pape, D., Harvey, N., Schurk
, R., Waterston, R., and Willson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com.
FEATURES
source
1..68
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl038-12"
/clone_lib="Gm-cl038"
/tissue_type="Senescing leaves, greenhouse grown"
/dev_stage="mature plants"
/lab_host="DH10B"
/notes="Vector: pT73Pac (pT73, Pharmacia); Site 1: EcoRI;
Site 2: HindIII; This cDNA library was constructed from
mRNA isolated from the senescing leaf tissue of mature
greenhouse grown plants of the cultivar Williams 82.
Complementary DNA was synthesized from mRNA using a 3'
anchored poly(dT) primer. EcoRI adapters were ligated to
the blunt ended cDNA fragments followed by digestion with
EcoRI and HindIII. The cDNA fragments were directionally
cloned into the EcoRI-HindIII restriction site of the
pT73-Pac vector. The ligated cDNA fragments were
transformed into DH10B host cells (Gibco BRL). This
library was constructed by Dr. Randy Shoemaker and Dr.
John Erpelting."
BASE COUNT 15 a 12 c 10 g 31 t
ORIGIN
Query Match 100.0%; Score 9; DB 164; Length 68;
Best Local Similarity 100.0%; Pred. No. 7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taattacc 9
|||||
Db 44 TAATATTAC 36

RESULT 14
A2755547/C
LOCUS A2755547 69 bp DNA GSS
DEFINITION ev01c06.x1 PAX3 CASTING Library 'ev' Homo sapiens genomic clone
A2755547
ACCESSION A2755547
VERSION A2755547
KEYWORDS GSS.
SOURCE human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 69)
Barber, T.D., Barber, M.C., Tomescu, O., Barr, F., Ruben, S. and
Friedman, T.B.
Cyclic amplification and selection of target genes regulated by
Pax3 and PAX3/FKHR in embryogenesis and alveolar rhabdomyosarcoma
Unpublished (2000)
Contact: Friedman TB
Laboratory of Molecular Genetics
National Institute on Deafness and Other Communication Disorders,
National Institutes of Health
5 Research Court, Room 2A-15, Rockville, MD 20850, USA
Tel: 301 402 7580
Fax: 301 496 7882
Email: friedman@nidcd.nih.gov
Plate: 01 row: c column: 06
Seq primer: -21M13 forward primer (ABI)
Class: random plasmid subclone.
Location/Qualifiers
1..69
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ev01c06"
/clone_lib="PAX3 CASTING Library 'ev'"
/sex="Male"
/lab_host="DH10B"
/notes="Vector: pGEM-T Easy; Human genomic DNA was
partially digested with Sau3AI, ligated to ds linkers,
and enriched for binding to human PAX3dQ+ protein using a
Whole Genome PCR-based strategy. DNA fragments containing
putative PAX3dQ+ binding sites were amplified by PCR and
cloned into pGEM-T Easy (Promega). The ligation products
were transformed into DH10B electrocompetent cells (Life
Technologies)."
BASE COUNT 18 a 13 c 11 g 27 t
ORIGIN
Query Match 100.0%; Score 9; DB 249; Length 69;
Best Local Similarity 100.0%; Pred. No. 7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taattacc 9
|||||
Db 48 TAATATTAC 40

RESULT 15
BG226871
LOCUS BG226871 71 bp mRNA EST
DEFINITION kp99a04.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5', mRNA
sequence.
ACCESSION BG226871
VERSION BG226871.1 GI:12714426
KEYWORDS EST.
SOURCE Strongyloides stercoralis.
ORGANISM Strongyloides stercoralis
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Strongyloides.
1 (bases 1 to 71)
McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
, Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
, Gibbons, M., Rittner, E., Bennett, J., Franklin, C., Tsagaris, R.,
, Ronko, L., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe
, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
, Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999

```

Thu Oct 25 13:08:48 2001

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIAID, NIH (tnutman@nih.gov). DNA sequencing by: Washington
University Genome Sequencing Center St. Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 62.

FEATURES

Location/Qualifiers
1. .71
/organism="Strongyloides stercoralis"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/clone_lib="MBN95TW-SSR"
/lab_host="XL-1 Blue MRF" (Stratagene)
/note="vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
rhabditiform larvae which had been isolated from gerbils
experimentally infected with larvae originally isolated
from experimentally infected dogs. cDNA was constructed
and using adaptors, was cloned unidirectionally into the
vector from the EcoRI site to the XhoI site. The library
has an unamplified titer of 1 x 10E5 pfu/ml and an
amplified, undiluted titer of 9 x 10E11 pfu/ml. The
average insert size of the unamplified library is 675 bp
(range, 100-1700)."
35 a 9 c 6 g 21 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 9; DB 174; Length 71;
Best Local Similarity 100.0%; Pred. No. 7e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 taatattac 9
| | | | | | | |
Db 62 TAATATTAC 70

Search completed: October 24, 2001, 13:13:56
Job time: 11821 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:30 ; Search time 180.6 seconds
(without alignments)
9,434 Million cell updates/sec

Title: US-09-462-955-3

Perfect score: 9

Sequence: 1 taatattac 9

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS-COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	100.0	17	1	US-08-058-023-3
2	9	100.0	20	1	US-08-297-703-3
3	9	100.0	20	1	US-08-229-145-1
4	9	100.0	20	1	US-08-229-145-6
5	9	100.0	20	3	US-09-392-350-36
6	9	100.0	20	5	PCT-US91-05802-1
7	9	100.0	20	5	PCT-US91-05802-6
8	9	100.0	21	1	US-08-229-145-10
9	9	100.0	21	3	US-09-046-578-11
10	9	100.0	21	5	PCT-US91-05802-10
11	9	100.0	34	4	US-08-793-701-45
12	9	100.0	35	3	US-09-051-914-10
13	9	100.0	37	2	US-08-805-191-7
14	9	100.0	72	1	US-08-433-126A-39
15	9	100.0	72	1	US-08-433-124A-39
16	9	100.0	72	3	US-08-976-413A-39
17	9	100.0	72	5	PCT-US96-06059-39
18	9	100.0	102	4	US-09-276-531-6
19	9	100.0	166	4	US-08-905-223-107
20	9	100.0	230	2	US-08-687-080-106
21	9	100.0	253	6	5242821-10
22	9	100.0	266	2	US-08-353-476-60
23	9	100.0	311	4	US-08-916-576B-40
24	9	100.0	311	4	US-08-916-576B-40
25	9	100.0	351	1	US-08-281-916-7
26	9	100.0	351	1	US-08-281-916-8
27	9	100.0	351	2	US-08-460-725-9

28	9	100.0	351	2	US-08-460-725-10	Sequence 10, Appl
29	9	100.0	358	1	US-08-686-878A-3	Sequence 3, Appl
30	9	100.0	399	1	US-07-885-970A-13	Sequence 13, Appl
31	9	100.0	399	1	US-08-298-587A-13	Sequence 13, Appl
32	9	100.0	399	1	US-08-530-797-12	Sequence 12, Appl
33	9	100.0	399	1	US-08-298-829-13	Sequence 12, Appl
34	9	100.0	399	2	US-08-787-335-12	Sequence 12, Appl
35	9	100.0	401	4	US-09-276-531-59	Sequence 59, Appl
36	9	100.0	405	1	US-08-307-499-45	Sequence 45, Appl
37	9	100.0	405	4	US-09-299-268-45	Sequence 45, Appl
38	9	100.0	485	3	US-09-284-782-17	Sequence 17, Appl
39	9	100.0	495	3	US-08-691-563C-59	Sequence 59, Appl
40	9	100.0	513	1	US-08-650-584-4	Sequence 4, Appl
41	9	100.0	513	3	US-09-060-584-4	Sequence 4, Appl
42	9	100.0	513	3	US-09-413-140A-4	Sequence 4, Appl
43	9	100.0	516	1	US-08-280-320-1	Sequence 1, Appl
44	9	100.0	551	4	US-08-998-416-86	Sequence 86, Appl
45	9	100.0	551	4	US-08-998-416-86	Sequence 86, Appl

ALIGNMENTS

RESULT 1
US-08-058-023-3
; Sequence 3, Application US/08058023
; Patent No. 5521302
; GENERAL INFORMATION:
; APPLICANT: Cook, Phillip D.
; TITLE OF INVENTION: OLIGONUCLEOTIDES HAVING CHIRAL
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk, 720 Kb
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/058,023
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Joseph Luccl
; REGISTRATION NUMBER: 33,307
; REFERENCE/DOCKET NUMBER: ISIS-1053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-058-023-3

Query Match 100.0%; Score 9; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9

Db 7 TAATATTAC 15
|||||

RESULT 2

US-08-297-703-3
; Sequence 3, Application US/08297703
; Patent No. 5506212
; GENERAL INFORMATION:
; APPLICANT: Hake, Glenn
; TITLE OF INVENTION: Stereoisomerically Pure
; TITLE OF INVENTION: Phosphorothioate
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John W. Caldwell
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/777,007
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Caldwell, John W.
; REGISTRATION NUMBER: 28,937
; REFERENCE/DOCKET NUMBER: ISIS-0015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-368-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-297-703-3

Query Match 100.0%; Score 9; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||

Db 10 TAATATTAC 18

RESULT 3

US-08-229-145-1
; Sequence 1, Application US/08229145
; Patent No. 5691461
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn D.
; TITLE OF INVENTION: INHIBITION OF CANDIDA USING
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 5691461ris
; STREET: One Liberty Place, 46th floor

CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/229,145
FILING DATE: Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: ISIS-1421
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: YES
US-08-229-145-1

Query Match 100.0%; Score 9; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
|||||

Db 10 TAATATTAC 18

RESULT 4

US-08-229-145-6
; Sequence 6, Application US/08229145
; Patent No. 5691461
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn D.
; TITLE OF INVENTION: INHIBITION OF CANDIDA USING
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 5691461ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/229,145
FILING DATE: Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: ISIS-1421
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-229-145-6

```

```

Query Match 100.0%; Score 9; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 taatattac 9
   |||||
DB 10 TAATATTAC 18

```

RESULT 5

```

US-09-392-350-36/c
; Sequence 36, Application US/09392350
; Patent No. 6133032
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF P13 KINASE P110 BETA EXPRESSION
; FILE REFERENCE: RFS-0075
; CURRENT APPLICATION NUMBER: US/09/392,350
; CURRENT FILING DATE: 1999-09-08
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 36
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-392-350-36

```

```

Query Match 100.0%; Score 9; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 taatattac 9
   |||||
DB 10 TAATATTAC 2

```

RESULT 6

```

PCT-US91-05802-1
; Sequence 1, Application PC/TUS9105802
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn C.
; APPLICANT: Ecker, David J.
; TITLE OF INVENTION: Inhibition of Candida
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESS: Norris
; STREET: One Liberty Place-- 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb
; MEDIUM TYPE: STORAGE
; APPLICANT: Hoke, Glenn C.
; APPLICANT: Ecker, David J.
; TITLE OF INVENTION: Inhibition of Candida
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESS: Norris
; STREET: One Liberty Place-- 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb
; MEDIUM TYPE: STORAGE
; COMPUTER: IBM PS/2

```

```

; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05802
; FILING DATE: 19910815
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Licata, Jane M.
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISIS-0432
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
PCT-US91-05802-1

```

```

Query Match 100.0%; Score 9; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 taatattac 9
   |||||
DB 10 TAATATTAC 18

```

RESULT 7

```

PCT-US91-05802-6
; Sequence 6, Application PC/TUS9105802
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn C.
; APPLICANT: Ecker, David J.
; TITLE OF INVENTION: Inhibition of Candida
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz
; ADDRESS: Norris
; STREET: One Liberty Place-- 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb
; MEDIUM TYPE: STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05802
; FILING DATE: 19910815
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Licata, Jane M.
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISIS-0432
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: Other nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
PCT-US91-05802-6

Query Match      100.0%; Score 9; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 10 TAATATTAC 18

RESULT 8
US-08-229-145-10
; Sequence 10, Application US/08229145
; Patent No. 5691461
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn D.
; APPLICANT: Ecker, David J.
; TITLE OF INVENTION: INHIBITION OF CANDIDA USING OLIGONUCLEOTIDES
; TITLE OF INVENTION: OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 5691461ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/229,145
; FILING DATE: Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: ISIS-1421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-229-145-10

Query Match      100.0%; Score 9; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 6 TAATATTAC 14

RESULT 9
US-09-046-578-11/c
; Sequence 11, Application US/09046578
; Patent No. 6087104
; GENERAL INFORMATION:
; APPLICANT: Yamada, Shoichi
; APPLICANT: Venkateswaran, Kasthuri
; APPLICANT: Ohashi, Ei-ji
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR DETECTION OF BACILLUS CEREUS GROUP BACTERIA HARMFUL TO MAMMALS, AND THE
; TITLE OF INVENTION: METHOD FOR DETECTION WITH THE OLIGONUCLEOTIDES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARMSTRONG, WESTERMAN, HATTORI, McLELAND &
; ADDRESSEE: NAUGHTON
; STREET: 1725 K Street, NW Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,578
; FILING DATE: 24-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeland, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 980313
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-659-2930
; TELEFAX: 202-887-0357
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-046-578-11

Query Match      100.0%; Score 9; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 13 TAATATTAC 5

RESULT 10
PCT-US91-05802-10
; Sequence 10, Application PC/TUS9105802
; GENERAL INFORMATION:
; APPLICANT: Hoke, Glenn C.
; APPLICANT: Ecker, David J.
; TITLE OF INVENTION: Inhibition of Candida
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place-- 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb
; MEDIUM TYPE: STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
```

```

; SOFTWARE: WORDPERFECT 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05802
; FILING DATE: 19910815
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Licata, Jane M.
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISIS-0432
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; HYPOTHEetical: NO
; ANTI-SENSE: YES
PCT-US91-05802-10

```

```

Query Match 100.0%; Score 9; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taatattac 9
Db 6 TAAATTATAC 14

```

```

RESULT 11
US-08-793-701-45/c
; Sequence 45, Application US/08793701
; Patent No. 6248581
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, Brigitte
; APPLICANT: LIM, Eng Mong
; APPLICANT: PORTNOI, Denis
; APPLICANT: BERTHET, Francois-Xavier
; APPLICANT: TIMM, Juliano
; TITLE OF INVENTION: MYCOBACTERIA FUNCTIONAL SCREENING AND/OR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: c/o FINNEGAN, HENDERSON, FARRABOW, GARRETT &
; ADDRESSEE: DUNNER, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,701
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR9501133
; FILING DATE: 30-AUG-1995
; APPLICATION DATA:
; APPLICATION NUMBER: FR 94/10585
; FILING DATE: 02-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonell, Leslie A.
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02356.0075

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4132
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
US-08-793-701-45

```

```

Query Match 100.0%; Score 9; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taatattac 9
Db 16 TAAATTATAC 8

```

```

RESULT 12
US-09-051-914-10
; Sequence 10, Application US/09051914A
; Patent No. 6096306
; GENERAL INFORMATION:
; APPLICANT: BRAVO, Alejandra
; APPLICANT: LERECLUS, Didier
; APPLICANT: SALAMITOU, Sylvie
; APPLICANT: SANCHIS, Vincent
; APPLICANT: AGAISSE, Herve
; TITLE OF INVENTION: NEW STRAINS OF BACILLUS THURINGIENSIS AND PESTICIDE
; FILE REFERENCE: 0660-0131-0X PCT
; CURRENT APPLICATION NUMBER: US/09/051,914A
; CURRENT FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: PCT/FR96/01684
; EARLIER FILING DATE: 1996-10-28
; EARLIER APPLICATION NUMBER: FR 95/12741
; EARLIER FILING DATE: 1995-10-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 35
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
US-09-051-914-10

```

```

Query Match 100.0%; Score 9; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 taatattac 9
Db 19 taatattac 27

```

```

RESULT 13
US-08-805-191-7
; Sequence 7, Application US/08805191
; Patent No. 5908924
; GENERAL INFORMATION:
; APPLICANT: Burdette, Douglas S
; APPLICANT: Zeikus, Joseph G
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THE GENE
; TITLE OF INVENTION: ENCODING THERMOANAEROBACTER ETHANOLICUS 39E
; TITLE OF INVENTION: SECONDARY-ALCOHOL DEHYDROGENASE AND ENZYME BIOCHEMICAL
; CHARACTERIZATION
; NUMBER OF SEQUENCES: 7

```

us-09-462-955-3.rni

Thu Oct 25 13:08:47 2001

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,191
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/012,331
FILING DATE: 27-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 660336.90608
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "primer"
US-08-805-191-7

Query Match 100.0%; Score 9; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 19 TAATATTAC 27

RESULT 14
US-08-433-126A-39
Sequence 39, Application US/08433126A
Patent No. 5688935
GENERAL INFORMATION:
APPLICANT: STEPHENS, ANDREW
APPLICANT: SCHNEIDER, DAN
TITLE OF INVENTION: NUCLEIC ACID LIGANDS OF TISSUE
TITLE OF INVENTION: TARGET
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM pc compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,126A
FILING DATE: 03 MAY 1995
CLASSIFICATION: 514

Query Match 100.0%; Score 9; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatattac 9
Db 43 TAATATTAC 51

RESULT 15
US-08-433-124A-39
Sequence 39, Application US/08433124A
Patent No. 5750342
GENERAL INFORMATION:
APPLICANT: STEPHENS, ANDREW
APPLICANT: SCHNEIDER, DAN
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: NUCLEIC ACID LIGANDS OF TISSUE
TITLE OF INVENTION: TARGET
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM pc compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,124A
FILING DATE: 03 MAY 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215

```

```

; REFERENCE/DOCKET NUMBER: NEX31.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-433-124A-39

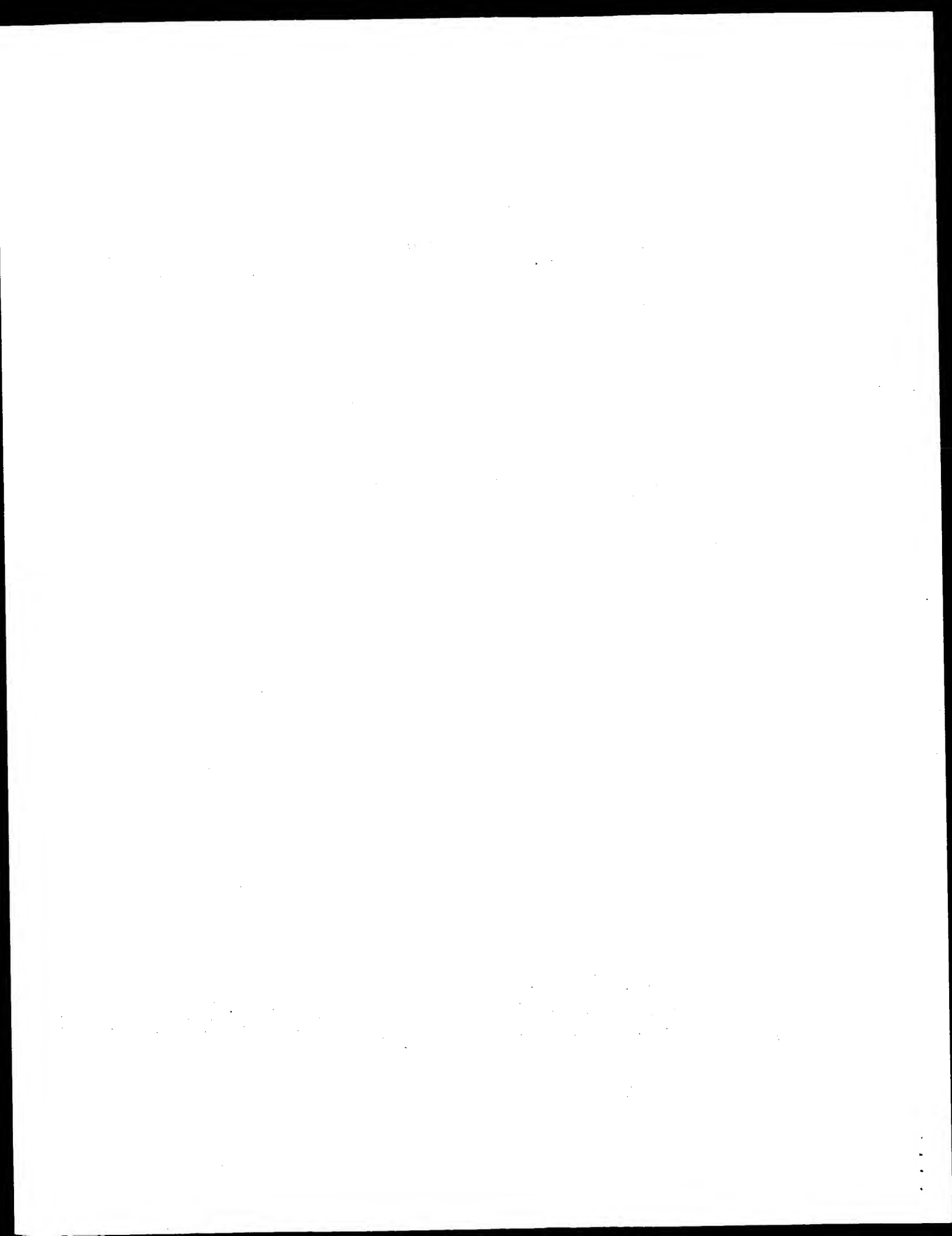
```

```

Query Match      100.0%; Score 9; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 taatattac 9
Db 43 TAATATTAC 51

```

Search completed: October 24, 2001, 10:00:31
Job time: 216 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:27 ; Search time 393.52 seconds
(without alignments)
14.360 Million cell updates/sec

Title: US-09-462-955-4

Perfect score: 9

Sequence: 1 taactactag 9

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601.*

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:	15:	16:	17:	18:	19:	20:	21:	22:
	/SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*	/SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	9	100.0	18	AAV33149	Plasmodium falciparum
C 2	9	100.0	18	AAV33146	Plasmodium vivax c
C 3	9	100.0	22	AAA30640	Human G protein-co
C 4	9	100.0	22	AAZ50574	3' PCR primer-1 to
C 5	9	100.0	22	AAF31603	Double-stranded DN
C 6	9	100.0	23	AAV45715	Human NKCC2 gene e
C 7	9	100.0	28	AAZ61141	Reverse PCR primer
C 8	9	100.0	30	AAZ87693	Human TPST-1 cDNA
C 9	9	100.0	30	AAZ99717	PCR primer for cDN
C 10	9	100.0	31	AAQ56057	Coconut Foliar Dec
C 11	9	100.0	31	AAZ02688	Gemini virus DNA f

C 12	9	100.0	31	20	AAZ02686	Gemini virus DNA f
C 13	9	100.0	32	18	AAZ90052	Primer for chicken
C 14	9	100.0	35	18	AAZ77100	Pisum sativum plas
C 15	9	100.0	35	21	AAA39333	Z. paucivorans 16S
C 16	9	100.0	36	20	AAZ28348	PCR primer PC641 f
C 17	9	100.0	38	18	AAZ62191	Primer PF668 for C
C 18	9	100.0	38	20	AAZ28335	PCR primer PF668 f
C 19	9	100.0	38	21	AAZ61496	Primer PCR2AS for
C 20	9	100.0	38	22	AAZ91511	Primer PCR2AS for
C 21	9	100.0	38	22	AAZ91519	CmR gene/PorA prom
C 22	9	100.0	39	21	AAA30642	Human G protein-co
C 23	9	100.0	39	21	AAZ50576	3' PCR primer-2 to
C 24	9	100.0	39	22	AAZ91503	Linker coding for
C 25	9	100.0	41	15	AAQ79998	CmR gene/PorA prom
C 26	9	100.0	53	21	AAA48234	Linker coding for
C 27	9	100.0	62	15	AAQ69204	T. reesei xylanase
C 28	9	100.0	67	21	AAA48233	T. reesei xylanase
C 29	9	100.0	74	22	AAZ31601	Variable oligonucle
C 30	9	100.0	93	16	AAZ26341	Human gene signatu
C 31	9	100.0	98	21	AAQ09195	Human secreted pro
C 32	9	100.0	101	16	AAQ97387	Human type i stero
C 33	9	100.0	110	18	AAV75998	Staphylococcus aur
C 34	9	100.0	112	21	AAZ17499	Human secreted pro
C 35	9	100.0	114	21	AAZ94806	Soybean microsate
C 36	9	100.0	116	21	AAZ10236	Human secreted pro
C 37	9	100.0	127	18	AAZ76841	Staphylococcus aur
C 38	9	100.0	138	21	AAZ18454	Human secreted pro
C 39	9	100.0	138	21	AAZ20546	Human secreted pro
C 40	9	100.0	149	21	AAA45738	Human secreted exp
C 41	9	100.0	151	21	AAZ26107	Human secreted pro
C 42	9	100.0	152	16	AAZ23458	Human gene signatu
C 43	9	100.0	159	21	AAA45500	Human secreted exp
C 44	9	100.0	165	21	AAZ22885	Human secreted pro
C 45	9	100.0	177	19	AAZ10779	Human biallelic po

ALIGNMENTS

RESULT 1
AAV33149/c
ID AAV33149 standard; DNA; 18 BP.
XX
AC AAV33149;
XX
DT 07-DEC-1998 (first entry)
XX
DE Plasmodium falciparum coxI gene PCR primer PFL.
XX
KW Malaria; infection; therapy; diagnosis; mitochondrion; coxI gene;
KW cytochrome C oxidase; PCR; primer; ss.
XX
OS Synthetic.
OS Plasmodium falciparum.
XX
PN WO9835057-A1.
XX
PD 13-AUG-1998.
XX
PF 05-FEB-1998; 98WO-IB00212.
XX
PR 26-SEP-1997; 97AU-0009481.
PR 06-FEB-1997; 97AU-0004953.
PR 21-APR-1997; 97AU-0006329.
XX
PA (MOLE-) INST MOLECULAR & CELL BIOLOGY.
PA (UYSI-) UNIV SINGAPORE NAT.
XX
PI Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;
XX WPI; 1998-447251/38.
XX
PT Detecting Plasmodium infection from hybridisation with

PT extrachromosomal element - providing genus or species specific
 PT diagnosis with few false negatives, in humans or animals
 XX
 XX Claim 6; Page 66; 120pp; English.

XX This synthetic oligonucleotide sequence (PFI) is derived from the
 CC Plasmodium falciparum mitochondrial cytochrome C oxidase (coxI)
 CC gene. In a claimed method, Plasmodium is detected in a human or
 CC animal sample (especially blood or dried blood) by treating it, or
 CC derived nucleic acid, with a Plasmodium extrachromosomal genetic
 CC element or derived nucleic acid (A) and detecting any hybridisation.
 CC (A) can include the PSI-PL470, PLH-PPH, PRB or PWQ gene (see
 CC AAV33135-38), the coxI gene or nucleic acids derived from them. Also
 CC new are (A)-specific probes and primers (see AAV33139-56). The method
 CC is used to diagnose Plasmodium infection. The high degree of
 CC similarity between (A) from different species allows development of
 CC genus- or species-specific assays that result in fewer false
 CC negatives than known methods (typically 1% against 3%). Primer PFI
 CC can be used with another primer (see AAV33145) for the species-
 CC specific detection of P. falciparum.

XX Sequence 18 BP; 4 A; 3 C; 3 G; 8 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;
 Best Local Similarity 100.0%; Pred. No. 3.8e-03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
 Db 10 TAATACTAG 2

RESULT 2
 ID AAV33146 standard; DNA; 18 BP.
 XX AAV33146;
 XX 07-DEC-1998 (first entry)
 XX Plasmodium vivax coxI gene PCR primer PV1.

XX Malaria; infection; therapy; diagnosis; mitochondrion; coxI gene;
 KW cytochrome C oxidase; PCR; primer; ss.

XX Synthetic.
 OS Plasmodium vivax.
 XX WO9835057-A1.

XX 13-AUG-1998.

XX 05-FEB-1998; 98WO-IB00212.

XX 26-SEP-1997; 97AU-0009481.

XX 06-FEB-1997; 97AU-0004953.

XX 21-APR-1997; 97AU-0006329.

XX (MOLE-) INST MOLECULAR & CELL BIOLOGY.

XX (UYSI-) UNIV SINGAPORE NAT.

XX Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;

XX WPI; 1998-447251/38.

XX Detecting Plasmodium infection from hybridisation with
 PT extrachromosomal element - providing genus or species specific
 PT diagnosis with few false negatives, in humans or animals
 XX

XX Claim 6; Page 65; 120pp; English.

XX This synthetic oligonucleotide sequence (PVI) is derived from the

CC Plasmodium vivax mitochondrial cytochrome C oxidase (coxI) gene.
 CC In a claimed method, Plasmodium is detected in a human or animal
 CC sample by treating it, or derived nucleic acid, with a Plasmodium
 CC extrachromosomal genetic element or derived nucleic acid (A) and
 CC detecting any hybridisation. (A) can include the PSI-PL470,
 CC PLH-PPH, PRB or PWQ gene (see AAV33135-38), the coxI gene or nucleic
 CC acids derived from them. Also new are (A)-specific probes and
 CC primers (see AAV33139-56). The method is used to diagnose Plasmodium
 CC infection. The sample can be blood or dried blood. The high
 CC degree of similarity between (A) from different species allows
 CC development of genus- or species-specific assays that result in
 CC fewer false negatives than known methods (typically 1% against 3%).
 CC Primer PVI can be used with another primer (see AAV33145) for the
 CC species-specific detection of P. vivax.

XX Sequence 18 BP; 3 A; 3 C; 3 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
 Db 10 TAATACTAG 2

RESULT 3
 ID AAA30640 standard; DNA; 22 BP.
 XX AAA30640;
 XX 21-AUG-2000 (first entry)

XX Human G protein-coupled receptor GHSR PCR primer, SEQ ID NO:84.
 DE G protein-coupled receptor; GPCR; constitutively active;
 KW intracellular loop 3; transmembrane domain 6; drug screening;
 KW agonist; antagonist; PCR primer; ss.
 XX Homo sapiens.
 OS WO200022129-A1.
 PN 20-APR-2000.
 XX 12-OCT-1999; 99WO-US23938.
 XX 13-OCT-1998; 98US-0170496.
 XX (AREN-) ARENA PHARM INC.
 XX Behan DP, Chalmers DT, Liaw CW;
 XX WPI; 2000-329165/28.

XX Non-endogenous constitutively activated human G protein-coupled
 PT receptors, useful for identifying agonists for use as pharmaceutical
 PT agents
 XX Example 1; Page 42; 341pp; English.
 XX The invention relates to constitutively active, non-endogenous versions
 CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
 CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
 CC and AAA30775-A30779). The mutant proteins of the invention contain a
 CC mutation in a portion of the protein comprising intracellular loop 3
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
 CC is substituted for an endogenous residue in IC3 at a position 16 amino
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence
 CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
 CC or Ala, and is preferably Lys. When the endogenous residue at this

position is Lys, this residue is replaced by His, Arg or preferably Ala. The 15 amino acid stretch between the substituted amino acid and the Pro may be endogenous, non-endogenous, or a mixture of endogenous and non-endogenous residues. The constitutively active GPCRs are useful for identifying antagonists, agonists and partial agonists for use as pharmaceutical agents. The mutant proteins are also useful in research settings for elucidating the roles of the receptors in normal and diseased conditions. Antagonists for a particular GPCR are useful for treating diseases and disorders associated with that receptor. Because the novel mutant GPCRs are constitutively active, they can be used directly for screening of compounds without the need for endogenous ligands. The present sequence represents a PCR primer used in an exemplification of the invention to isolate cDNA encoding a human wild-type GPCR for cloning, an exemplification of the invention. The GPCR cDNA was subjected to site-directed mutagenesis (SDM) to generate DNA encoding the corresponding mutant of the invention.

Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.8e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 taatactag 9
Db 9 taatactag 17

RESULT 4
AAZ50574
ID AAZ50574 standard; DNA; 22 BP.

XX AAZ50574;

XX 20-JUN-2000 (first entry)

XX 3' PCR primer-1 to amplify human GHSR DNA.

XX G protein-coupled orphan receptor; GPCR; agonist; G protein;
XX GPCR fusion protein; inverse agonist; drug; treatment; PCR primer;
XX GHSR; G protein-coupled receptor; human; ss.

OS Homo sapiens.

XX WO200006597-A2.

XX 10-FEB-2000.

XX 30-JUL-1999; 99WO-US17425.

XX 31-JUL-1998; 98US-0094879.

XX 30-OCT-1998; 98US-0106300.

XX 04-DEC-1998; 98US-0110906.

XX 26-FEB-1999; 99US-0121851.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw C, Lin I, Lowitz K, Chen R;

XX WPI; 2000-195260/17.

XX Identification of a compound useful as a therapeutic agent, comprises
XX identifying a compound against constitutively activated G
XX protein-coupled orphan receptors

XX Example 2; Page 27; 123pp; English.

XX The patent discloses a method of identifying agonists and inverse or
XX partial agonists to the endogenous, constitutively activated
XX G protein-coupled orphan receptors (GPCRs), by contacting them with a
XX GPCR fusion protein comprising a GPCR and a G protein. Determining
XX expression of GPCRs in tissue samples can be used to identify related

CC diseases. Inverse agonists to these receptors can be used as drugs for
CC treating GPCR-related diseases. The present sequence is a 3' PCR primer,
CC used in primary PCR reaction to amplify human GHSR DNA from
CC hippocampus cDNA template.

XX Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.8e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 taatactag 9
Db 9 taatactag 17

RESULT 5

AAF31603
ID AAF31603 standard; DNA; 22 BP.

XX AAF31603;

XX 09-APR-2001 (first entry)

XX Double-stranded DNA tag PCR primer pTag2.

XX attenuated microorganism; signature tagged transposon mutant;
XX mutant library; mycobacterial infection; actinomycetales;
XX antibacterial; immunostimulant; vaccine; PCR primer; ss.

OS Unidentified.

XX WO200102555-A1.

XX 11-JAN-2001.

XX 06-JUL-2000; 2000WO-IB00950.

XX 06-JUL-1999; 99US-0142982.

XX 08-JUL-1999; 99US-0142833.

XX (INSP) INST PASTEUR.

XX Gicquel B, Guilhot C, Camacho L;

XX WPI; 2001-091804/10.

XX Screening a mutant library for mutants unable to grow under specific
XX conditions and for identifying loci involved in pathogenicity,
XX comprises using signature tagged transposon mutagenesis

XX Example 4; Page 19; 159pp; English.

XX The present sequence is given in a specification relating to a method for
XX screening a library of mutants. The method comprises constructing a
XX library with insertions in genes and/or regulatory regions of the
XX organisms of interest, where the insertion contains a tag and/or a
XX transposon associated with a tag. The mutants are identified by
XX hybridisation of the tags to known sequences. The method is useful for
XX treating an individual suffering from a mycobacterial infection,
XX suspected of being infected with a Mycobacterium, or having been
XX exposed to an infectious Mycobacterium. It is also useful for
XX identifying and isolating mutants of actinomycetales and for identifying
XX mutants of microorganisms, preferably an actinomycetales, such as
XX M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
XX M. paratuberculosis, that is unable to grow under specific conditions.
XX It is especially useful for identifying loci involved in pathogenicity.
XX It is useful in constructing vaccines. The method can be used to screen
XX multiple libraries concurrently. It can screen libraries of different
XX organisms or different strains of the same organism. The present
XX sequence was used to create the tag used in the construction of tagged

Thu Oct 25 13:08:51 2001

us-09-462-955-4.rng

CC mutants.
XX Sequence 22 BP; 7 A; 5 C; 1 G; 9 T; 0 other;
SQ

Query Match 100.0%; Score 9; DB 22; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.8e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taactactag 9
DB 3 taactactag 11

RESULT 6
AAV45715
ID AAV45715 standard; DNA; 23 BP.
XX
AC AAV45715;
XX
DT 21-DEC-1998 (first entry)
XX
DE Human NKCC2 gene exon 19 forward primer hNKCC2ex19.
XX
KW Na-K-2Cl cotransporter; NKCC2; human; Bartter's syndrome;
KW ion transport; hypokalaemic alkalosis; hypercalciuria;
KW nephrocalcinosis; diagnosis; therapy; SSCP; primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9829431-A1.
XX
PD 09-JUL-1998.
XX
PF 19-DEC-1997; 97WO-US23553.
XX
PR 31-DEC-1996; 96US-0778052.
XX
PA (UYUA) UNIV YALE.
XX
PI Lifton RP, Simon DB;
PS WPI; 1998-388029/33.
DR
XX
XX Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2Cl
PT cotransporter genes - useful for developing products for the
PT diagnosis and treatment of ion transport disorders, e.g. Gitelman's
PT Syndrome or Bartter's Syndrome
XX
XX Example 2; Page 65; 105pp; English.

XX Primers hNKCC2ex19 forward and reverse (see AAV45715 and AAV45716,
CC respectively) are designed to amplify exon 19 of the human NKCC2
CC gene (see AAV4562) that codes for Na-K-2Cl cotransporter NKCC2
CC protein (see AAV23683). Both primers are located within introns of
CC the gene. 27 Sets of specific primers (see AAV45677-V45730) were
CC used for SSCP analysis of NKCC2. Amplified products were analysed
CC for molecular variants by electrophoresis, and identified variants
CC were sequenced. Complete linkage of Bartter's syndrome with NKCC2
CC was demonstrated. Identification of the molecular basis of
CC Bartter's syndrome allows for the genetic diagnosis of this
CC disorder. The invention provides products and methods useful for
CC diagnosis and treatment of Bartter's syndrome and other ion
CC transport disorders.
XX Sequence 23 BP; 8 A; 4 C; 5 G; 6 T; 0 other;
SQ

Query Match 100.0%; Score 9; DB 19; Length 23;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taactactag 9
DB 2 taactactag 10

RESULT 7
AAZ61141/C
ID AAZ61141 standard; DNA; 28 BP.
XX
AC AAZ61141;
XX
DT 30-MAY-2000 (first entry)
XX
DE Reverse PCR primer used to amplify DNA encoding the Shiga toxin.
XX
KW Chemokine receptor; ligand; inflammatory response; immune effector cell;
KW secondary tissue damage; central nervous system injury; Shiga;
KW CNS inflammatory disease; neurodegenerative disorder; heart disease;
KW inflammatory eye disease; inflammatory bowel disease; PCR primer;
KW inflammatory joint disease; inflammatory kidney; renal disease;
KW inflammatory lung disease; inflammatory nasal disease; thyroiditis;
KW inflammatory thyroid disease; cytokine-regulated cancer; ss.
XX
OS Shigella dysenteriae.
XX
PN WO200004926-A2.
XX
PD 03-FEB-2000.
XX
PF 21-JUL-1999; 99WO-CA00659.
XX
PR 22-JUL-1998; 98US-0120523.
XX
PA (OSPR-) OSPREY PHARM LTD.
XX
PI McDonald JR, Coggins PJ;
XX
DR WPI; 2000-182542/16.
XX
PT A new therapeutic agent comprising a conjugate for treating secondary
PT tissue damage and other disease conditions like Alzheimer's disease,
PT stroke, Parkinson's disease and atherosclerosis
XX
PS Example 1; Page 135; 204pp; English.

XX PCR primers AAZ61140-42 were used to amplify nucleic acids encoding
CC the Shiga toxin. The toxin can be incorporated into the conjugates
CC of the invention. The specification describes a conjugate, comprising
CC a targeted agent and a chemokine receptor ligand. The conjugate
CC binds to a chemokine receptor resulting in internalisation of the
CC targeted agent in cells bearing the receptor. The conjugates are
CC used for formulating a medicament or for treating disorders associated
CC with inflammatory responses resulting from activation, proliferation
CC and migration of immune effector cells. The disorders or disease states
CC comprise secondary tissue damage such as central nervous system (CNS)
CC injury, CNS inflammatory diseases, neurodegenerative disorders, heart
CC disease, inflammatory eye diseases, inflammatory bowel diseases,
CC inflammatory joint diseases, inflammatory kidney or renal diseases,
CC inflammatory lung diseases, inflammatory nasal diseases, inflammatory
CC thyroid disease such as thyroiditis, or cytokine-regulated cancers.
XX Sequence 28 BP; 6 A; 7 C; 8 G; 7 T; 0 other;
SQ

Query Match 100.0%; Score 9; DB 21; Length 28;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 taactactag 9
DB 19 TAATACTAG 11

```

RESULT 8
AAZ87693/c
ID AAX87693 standard; DNA: 30 BP.
XX
AC AAX87693;
XX
DT 26-OCT-1999 (first entry)
XX
DE Human TPST-1 cDNA bottom strand PCR primer.
XX
TPST-1; tyrosylprotein sulfotransferase; tyrosine O-sulfation;
KW post-translation modification; human; PCR; primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9938980-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01756.
XX
PR 09-SEP-1998; 98US-0150141.
PR 29-JAN-1998; 98US-0072994.
PR 09-SEP-1998; 98US-0150133.
XX
PA (OKLA ) UNIV OKLAHOMA STATE.
XX
PI Moore KL;
XX
DR WPI; 1999-494094/41.
XX
PT Human, murine and Caenorhabditis elegans tyrosylprotein
PT sulfotransferases, useful for post-translational tyrosine sulfation
XX
PS Example; Page 39; 123pp; English.
XX
CC This bottom strand primer was used with a top strand primer (see
CC AAX87692) in the PCR amplification of the human tyrosylprotein
CC sulfotransferase (TPST-1) coding sequence (see AAX87686) using EST
CC clone #116978 as template. The primers introduce a 5' BamHI site
CC and a 3' SpeI site. The amplified cDNA was ligated into vector
CC pCDNA3.1(+), and TPST-1 (see AAY06623) was expressed as a fusion
CC protein containing an N-terminal epitope for HPCA (see AAY06630)
CC in mammalian cells. TPST-1 is useful for the post-translational
CC tyrosine O-sulfation of proteins and peptides.
XX
SQ Sequence 30 BP; 5 A; 9 C; 6 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taactactag 9
Db 12 TAATACTAG 4

RESULT 9
AAZ99717/c
ID AAZ99717 standard; DNA: 30 BP.
XX
AC AAZ99717;
XX
DT 12-JUL-2000 (first entry)
XX
DE PCR primer for cDNA encoding human tyrosylprotein sulfotransferase 1.
XX
KW Human; tyrosylprotein sulfotransferase 1; TPST-1; tyrosine O-sulfation;
KW P-selectin glycoprotein ligand-1; PSGL-1; anti-inflammatory;
KW neutrophil binding; PCR primer; ss.
XX

```

```

OS Homo sapiens.
XX WO200014250-A1.
XX
PD 16-MAR-2000.
XX
PF 23-JUL-1999; 99WO-US16750.
XX
PR 09-SEP-1998; 98US-0150133.
PR 09-SEP-1998; 98US-0150141.
PR 28-JAN-1999; 99WO-US01756.
XX
PA (OKLA ) UNIV OKLAHOMA STATE.
XX
PI Moore KL;
XX
DR WPI; 2000-271057/23.
XX
PT New human or murine tyrosylprotein sulfotransferase, used for in vitro
PT or in vivo sulfation of proteins and to screen for anti-inflammatory
PT agents.
XX
PS Example; Page 40; 141pp; English.
XX
CC PCR primers AAZ99716-17 were used to amplify cDNA encoding human
CC tyrosylprotein sulfotransferase 1 (TPST-1) polypeptide. TPST-1 catalyses
CC the transfer of sulfate from 3'-phosphoadenosine-5'-phosphosulfate to
CC tyrosine residues (tyrosine O-sulfation) in proteins. TPST-1
CC polypeptides are used for in vitro or in vivo sulfation of proteins or
CC peptides, particularly P-selectin glycoprotein ligand-1 (PSGL-1) or its
CC functional fragments. They are also used to raise specific antibodies,
CC and are useful as immunoassay reagents, and to identify specific
CC inhibitors. These inhibitors are potentially useful as anti-inflammatory
CC agents, by inhibiting PSGL-1 mediated binding of neutrophils to
CC endothelial cells. The TPST-1 polynucleotide is useful for identifying
CC polymorphisms that may be associated with a disease phenotype.
XX
SQ Sequence 30 BP; 5 A; 9 C; 6 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taactactag 9
Db 12 TAATACTAG 4

RESULT 10
AAQ56057
ID AAQ56057 standard; DNA: 31 BP.
XX
AC AAQ56057;
XX
DT 12-AUG-1994 (first entry)
XX
DE Coconut Foliar Decay Virus promoter stem-loop.
XX
KW Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;
KW phloem-specific; stem-loop structure; transgenic plant; ds.
XX
OS Coconut Foliar Decay Virus.
XX
FH Key Location/Qualifiers
FH stem_loop 1..31
FT /*tag= a
FT /function= promoter
FT /note= "loop has homology to geminivirus sequence"
XX
PN DE4306832-C.
XX
PD 24-FEB-1994.

```

Thu Oct 25 13:08:51 2001

```

XX PF 04-MAR-1993; 93DE-4306832.
XX PF 04-MAR-1993; 93DE-4306832.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1994-058406/08.
XX DR Use of coconut foliar decay virus DNA as promoter - for
XX PT tissue-specific gene expression in transgenic plants
XX PT Claim 1; Fig 2; 8pp; German.
XX PS A DNA fragment from the CFDV genome can be used as a phloem-specific
XX CC promoter in the construction of transgenic plants. The promoter is
XX CC strong; it has 30-50% of the activity of the CaMV 35S promoter in
XX CC tobacco plants.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
Db 12 taatactag 20

RESULT 11
AAX02688
ID AAX02688 standard; DNA; 31 BP.
AC AAX02688;
XX 10-MAY-1999 (first entry)
XX Gemini virus DNA fragment stem loop.
XX Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX OS Gemini virus.
XX Key Location/Qualifiers
XX FT stem_loop 1..31
XX FT /*tag= a
XX FT misc_binding 1..11
XX FT /*tag= b
XX FT /note= "Region binds to nucleotides 21 to 31"
XX FT misc_binding 21..31
XX FT /*tag= c
XX FT /note= "Region binds to nucleotides 1 to 11"
XX DE19730502-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030502.
XX PF 16-JUL-1997; 97DE-1030502.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096863/09.
XX CC Coconut foliar decay virus promoters - for gene expression in

```

```

PT bacteria and yeasts
XX Disclosure; Fig 2; 14pp; German.
XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX that includes the stem-loop structure of CFDV DNA but lacks the
XX translation start codons of open reading frames ORF1 and/or ORF2. The
XX new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX promoter for tissue-specific (especially phloem-specific) gene expression
XX in plants and for production of chimeric constructs for transient or
XX stable expression. Certain fragments of CFDV DNA have stronger promoter
XX activity in E. coli than the CaMV 35S promoter.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
Db 12 taatactag 20

RESULT 12
AAX02686
ID AAX02686 standard; DNA; 31 BP.
AC AAX02686;
XX 10-MAY-1999 (first entry)
XX Gemini virus DNA fragment stem loop.
XX Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX OS Gemini virus.
XX Key Location/Qualifiers
XX FT stem_loop 1..31
XX FT /*tag= a
XX FT misc_binding 1..11
XX FT /*tag= b
XX FT /note= "Region binds to nucleotides 21 to 31"
XX FT misc_binding 21..31
XX FT /*tag= c
XX FT /note= "Region binds to nucleotides 1 to 11"
XX DE19730535-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030535.
XX PF 16-JUL-1997; 97DE-1030535.
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096867/09.
XX DR Coconut foliar decay virus promoters - for gene expression in
XX PT plants, bacteria and yeasts
XX PS Disclosure; Fig 2; 12pp; German.
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX that includes the stem-loop structure of CFDV DNA but lacks the
XX translation start codons of open reading frames ORF1 and/or ORF2. The
XX new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a
XX promoter for tissue-specific (especially phloem-specific) gene expression
XX in plants and for production of chimeric constructs for transient or
XX stable expression. Certain fragments of CFDV DNA have stronger promoter
XX activity in E. coli than the CaMV 35S promoter.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

```

CC promoter for tissue-specific (especially phloem-specific) gene expression
 CC in plants and for production of chimeric constructs for transient or
 CC stable expression. Certain fragments of CPDV DNA have stronger promoter
 CC activity in *E. coli* than the CaMV 35S promoter.
 XX
 SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 taatactag 9
 |||||
 Db 12 taatactag 20

RESULT 13
 AAT90052/c
 ID AAT90052 standard; DNA; 32 BP.
 AC AAT90052;
 XX
 DT 27-MAR-1998 (first entry)
 XX
 DE Primer for chicken urokinase cDNA.
 XX
 KW Low molecular weight; chicken; urokinase; reporter; screening;
 KW PCR primer; ss.
 XX
 OS Gallus domesticus.
 XX
 PN WC9734144-A1.
 XX
 PD 18-SEP-1997.
 XX
 PF 11-MAR-1997; 97WO-EP01219.
 XX
 PR 20-DEC-1996; 96DE-4054616.
 PR 14-MAR-1996; 96DE-4011780.
 XX
 PA (SCHD) SCHERING AG.
 XX
 PI Langer G, Schleuning W, Toschi L;
 XX
 DR WPI; 1997-470977/43.
 XX

Determining effect of test substance on cultured cells - by
 transfecting cells with plasmid having gene for low molecular weight
 hen urokinase as reporter, which can be detected by plasminogen
 activation and colour forming or fluorochrome substrate reaction
 XX
 PS Example 1; Page 16; 32pp; German.

The present sequence was used in the development of a method for
 determining the effect of test substance in a cell culture system.
 The method comprises preparing a plasmid having an expression
 cassette including the low molecular weight gene for chicken
 urokinase as a reporter, transfecting eukaryotic cells
 with the plasmid, inducing the cells with the test substance,
 harvesting the cell supernatant and detecting urokinase by adding
 chicken plasminogen (this addition precisely defines the start of
 the reaction), so that this is converted to plasmin by the
 urokinase, and detecting the plasmin in the supernatant with a
 highly sensitive chromogenic or fluorochromic substrate. The method
 can be used to screen new compounds, non-steroidal glucocorticoids
 or anti-glucocorticoids, determine receptors present in cells,
 particularly steroid hormone receptors and investigate promoter
 dependent modulated transcription processes. The method does not
 require radioisotopes or extraction of reporter gene products, is
 sensitive, non-destructive and can be performed in microtitre
 plates. Urokinase is small (so easily separated from endogenous
 proteases by gel electrophoresis), and unlike other reporter gene

CC products is stable in culture supernatant, permitting direct
 CC measurement. The method can be calibrated using a precise number of
 CC transfected cells, avoiding difficulties of calibration based on
 CC enzymatic activity. The urokinase/plasminogen reaction represents a
 CC long lived amplification system, which is well characterised.
 XX
 SQ Sequence 32 BP; 5 A; 7 C; 11 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 18; Length 32;
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 taatactag 9
 |||||
 Db 20 TAATACTAG 12

RESULT 14
 AAT77100/c
 ID AAT77100 standard; DNA; 35 BP.
 XX
 AC AAT77100;
 XX
 DT 20-JAN-1998 (first entry)
 XX
 DE Pisum sativum plastocyanin promoter sub-sequence.
 XX
 KW plastocyanin; promoter; enhancer; reporter gene; pea;
 KW transgenic plant; ds.
 XX
 OS Pisum sativum.
 XX
 FH Key Location/Qualifiers
 FT enhancer 1..31
 FT /*tag= a
 FT misc_feature 1..4
 FT /*tag= b
 FT /label= sticky_end
 FT /note= "The 5' end of complementary strand
 FT overhangs the 3' end of this strand by the
 FT sequence 5'-GATC-3'."
 XX
 PN WO9720056-A2.
 XX
 PD 05-JUN-1997.
 XX
 PF 26-NOV-1996; 96WO-GB02910.
 XX
 PR 29-NOV-1995; 95GB-0024350.
 XX
 PA (ADTE-) ADVANCED TECHNOLOGIES CAMBRIDGE.
 XX
 PI Gray JC, Sandhu JS, Webster CI;
 XX
 DR WPI; 1997-310610/28.
 XX
 PT Enhancer from pea plastocyanin promoter region - used to produce
 XX transgenic plants having increased expression of a desired gene
 PS Claim 11; Figure 7; 46pp; English.

This sequence represents the pea plastocyanin promoter enhancer
 which is a sub-sequence of AAT77099. This enhancer has been used to
 increase the expression of the GUS reporter gene in tobacco where the
 GUS gene was under the control of the pEF promoter and 35S CaMV
 promoter. The enhancer was also used joined to the -330 to +1 region of
 the patatin promoter pS20 which was joined to the GUS reporter and
 transformed into potato using A. tumefaciens. Expression of the gene
 promoter of one or more genes of a plant increases in green or non-green
 tissues, especially roots, tubers, seeds, flowers or leaves and
 transformed plants may be dicotyledonous or monocotyledonous. This
 enhancer may operate both in normal and reverse directions and can be

Thu Oct 25 13:08:51 2001

us-09-462-955-4.rng

CC attached to the promoter or terminator of the gene to be expressed.

XX
SQ Sequence 35 BP; 17 A; 4 C; 2 G; 12 T; 0 other;

Query Match 100.0%; Score 9; DB 18; Length 35;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
| | | | |
Db 19 TAATACTAG 11

RESULT 15
AAA39333/C
ID AAA39333 standard; DNA; 35 BP.
XX
AC AAA39333;
XX
DT 11-SEP-2000 (first entry)
XX
DE Z. paucivorans 16S rRNA and 23S rRNA gene spacer region SEQ ID NO:8.
XX
DE 16S rRNA; 23S rRNA; spacer region; detection; microbe; beer;
KW brewery; ds.
XX
OS Zymophilus paucivorans.
XX
PN JP2000106881-A.
XX
PD 18-APR-2000.
XX
PF 08-OCT-1998; 98JP-0286697.
XX
PR 08-OCT-1998; 98JP-0286697.
XX
PA (ASAK) ASAHI BREWERIES LTD.
XX
DR WPI; 2000-353477/31.
XX
PT A gene sequence spacer region between a gene encoding 16S rRNA and a
gene encoding 23S rRNA, useful for detection of microbes -
XX
PS Claim 8; Page 2; 18pp; Japanese.
XX
CC The present invention describes gene sequences of spacer region between
a gene encoding 16S rRNA and a gene encoding 23S rRNA for the detection
of microbes. Also described is a method for the detection of microbes
in which an oligonucleotide prepared from a spacer region as described
above is made to function as a primer for nucleic acid synthesis and
treated by gene amplification. The method is used for the detection of
microbes found in beer breweries. The present sequence represents a
region DNA sequence, from the present invention.
XX
SQ Sequence 35 BP; 11 A; 3 C; 7 G; 14 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 35;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
| | | | |
Db 15 TAATACTAG 7

Search completed: October 24, 2001, 10:07:28
Job time: 633 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:56 ; Search time 5479.82 Seconds
(without alignments)
15.525 Million cell updates/sec

Title: US-09-462-955-4
Perfect score: 9
Sequence: 1 taactagtag 9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estomi.*
69: em_estov1.*
70: em_estov2.*
71: em_estov3.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est33.*
111: gb_est34.*
112: gb_est35.*
113: gb_est36.*
114: gb_est37.*
115: gb_est38.*
116: gb_est39.*

Thu Oct 25 13:08:55 2001

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	9	100.0	40	167	BE386120	BE386120 601277233
C 2	9	100.0	55	249	AZ769245	AZ769245 1M0569010
C 3	9	100.0	57	14	AF097417	AF097417 AF097417
C 4	9	100.0	57	243	A2434302	A2434302 1M0220E09
C 5	9	100.0	58	10	AA637860	AA637860 vr29f01.r
C 6	9	100.0	59	113	AW215812	AW215812 uc99e06.x
C 7	9	100.0	59	113	AW215820	AW215820 uc99f06.x
C 8	9	100.0	62	258	TAB2F10Q	AL459971 T. brucei
C 9	9	100.0	66	250	AZ808515	AZ808515 2M0072A07
C 10	9	100.0	67	24	A1761414	A1761414 w65606.x
C 11	9	100.0	71	4	AA250291	AA250291 mz61502.r
C 12	9	100.0	71	123	AW989354	AW989354 ut25b07.y
C 13	9	100.0	71	138	BE627067	BE627067 uu15f06.y
C 14	9	100.0	71	138	BE630111	BE630111 uu15f06.x
C 15	9	100.0	71	138	BE688187	BE688187 uw43b06.x
C 16	9	100.0	71	138	BE692587	BE692587 uw43b06.y
C 17	9	100.0	71	141	BE849664	BE849664 uw87f04.y
C 18	9	100.0	73	167	BE448984	BE448984 ut50h06.y
C 19	9	100.0	75	2	AA079298	AA079298 zm97g12.s
C 20	9	100.0	78	113	AW226900	AW226900 um63c01.y
C 21	9	100.0	80	138	BE627009	BE627009 uu08h09.y
C 22	9	100.0	84	2	AA079297	AA079297 zm97g12.r
C 23	9	100.0	84	106	AA012912	AA012912 AU012912
C 24	9	100.0	85	13	AA937702	AA937702 o304g12.s
C 25	9	100.0	87	11	AA760419	AA760419 vv76h02.r
C 26	9	100.0	87	249	AZ783472	AZ783472 2M0025B18
C 27	9	100.0	90	155	C01576	C01576 HUMG000858
C 28	9	100.0	93	159	N28044	N28044 EST000258.S
C 29	9	100.0	94	20	AI465178	AI465178 mv98a02.x
C 30	9	100.0	94	241	AZ309606	AZ309606 1M0016B06
C 31	9	100.0	94	247	AZ660200	AZ660200 1M0538108
C 32	9	100.0	95	31	AV562434	AV562434 AV562434
C 33	9	100.0	97	7	AA419984	AA419984 vf38h02.r
C 34	9	100.0	99	107	AU014567	AU014567 AU014567
C 35	9	100.0	100	10	AA689650	AA689650 vs10g11.r
C 36	9	100.0	100	24	AI718783	AI718783 as58h08.x
C 37	9	100.0	101	164	BE152001	BE152001 QV1-HT031
C 38	9	100.0	101	164	BE179567	BE179567 IL3-HT061
C 39	9	100.0	102	166	BE365112	BE365112 P11_24_B0
C 40	9	100.0	102	189	T61718	T61718 yb92e07.r1
C 41	9	100.0	102	237	AZ059959	AZ059959 RPCT-23-4
C 42	9	100.0	104	8	AA518125	AA518125 v123d11.f
C 43	9	100.0	104	121	AW810470	AW810470 MR4-ST012
C 44	9	100.0	104	164	BE145670	BE145670 IL0-HT020
C 45	9	100.0	105	1	AA014683	AA014683 mh09g05.f

ALIGNMENTS

RESULT 1	
BE386120/c	
LOCUS	40 bp mRNA
DEFINITION	601277233F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3617961 5', mRNA sequence.
ACCESSION	BE386120
VERSION	BE386120.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 40)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC/DCID/BTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM286 row: b column: 10.

FEATURES

Location/Qualifiers
1..40
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3617961"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5'-adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 12 a 8 c 11 g 9 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Score 9; DB 167; Length 40;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9

|||||
Db 19 TAATACTAG 11

RESULT 2

AZ769245 55 bp DNA GSS 16-FEB-2001
LOCUS IM0569010R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0569010 R, DNA sequence.
ACCESSION AZ769245
VERSION AZ769245.1 GI:12889180
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 55)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0369 row: 0 column: 10
Seq primer: CACACAGGAACACATGACC
Class: plasmid ends
High quality sequence stop: 55.
Location/Qualifiers
1..55
/organism="Mus musculus"

FEATURES

Location/Qualifiers
1..55
/organism="Mus musculus"

```
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0569010"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
33 a 4 c 4 g 14 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 9; DB 249; Length 55;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 10 TAATACTAG 18

RESULT 3
AF097417 57 bp mRNA EST 10-NOV-1999
LOCUS AF097417 Vespertilio superans library Vespertilio superans cDNA, mRNA sequence.
DEFINITION AF097417
ACCESSION AF097417.1 GI:6341356
VERSION AF097417.1
KEYWORDS EST.
SOURCE Asian particolored bat.
ORGANISM Vespertilio superans
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Chiroptera; Microchiroptera; Vespertilionidae; Vespertilio.
1 (bases 1 to 57)
REFERENCE 1 (bases 1 to 57)
AUTHORS Kang,H.S. and Kim,J.W.
TITLE Differential Gene Expression of Active and Hibernating States in Bats
JOURNAL Unpublished (1999)
COMMENT Contact: Kang HS
Anatomy
Wonju College of Medicine, Yonsei University
Ilsandong 162, Wonju, Kangwon 220-701, Korea
Email: mhkim@yunc.yonsei.ac.kr.
FEATURES
Location/Qualifiers
1..57
/organism="Vespertilio superans"
/db_xref="taxon:105273"
/clone_lib="Vespertilio superans library"
/note="country=South Korea; Kangwon Province"
19 a 9 c 20 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 9; DB 14; Length 57;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0569010"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
33 a 4 c 4 g 14 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 9; DB 243; Length 57;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 42 TAATACTAG 50

RESULT 4
AF097417 57 bp DNA GSS 03-OCT-2000
LOCUS AF097417 Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0220509 R, DNA sequence.
DEFINITION AF097417
ACCESSION AF097417.1 GI:10558315
VERSION AF097417.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 57)
REFERENCE 1 (bases 1 to 57)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0220 row: E column: 09
Seq primer: CACACAGGAACACGTCATGACC
Class: plasmid ends
High quality sequence stop: 57.
Location/Qualifiers
1..57
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0220509"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
12 a 14 c 9 g 22 t
BASE COUNT
ORIGIN
```


Thu Oct 25 13:08:55 2001

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps@email.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

MGI:1031143
 Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
 1..59
 /organism="Mus musculus"
 /strain="CZECH II"
 /db_xref="taxon:10090"
 /clone="IMAGE:2650691"
 /clone_lib="NCI_CGAP_Lu30"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DHL08"
 /note="Organ: lung; Vector: PCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; transgenic model WNT-1, expression driven by
 MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
 dt. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 14 a 11 c 7 g 27 t

Query Match 100.0%; Score 9; DB 113; Length 59;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
 |||||
Db 13 TAATACTAG 5

RESULT 8
TA82F10Q 62 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 82f10, reverse sequence,
DEFINITION genomic survey sequence.
ACCESSION AL459971
VERSION AL459971.1 GI:11860296
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.

REFERENCE 1 (bases 1 to 62)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
 Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
 Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
JOURNAL Project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 nhlesanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
 Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (FRE927/4 Cutat 10.1) was mechanically sheared
 to give a tight size distribution (800-1000 bp). The v + i method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES Location/Qualifiers
 1..62
 /organism="Trypanosoma brucei"
 /strain="FRE927"
 /db_xref="taxon:5691"
 /clone="82f10"

BASE COUNT 27 a 8 c 11 g 16 t

ORIGIN

Query Match 100.0%; Score 9; DB 258; Length 62;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
 |||||
Db 10 TAATACTAG 18

RESULT 9
AZ808515/c 66 bp DNA GSS 20-FEB-2001
LOCUS 2M0072A07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0072A07 F, DNA sequence.

ACCESSION AZ808515
VERSION AZ808515.1 GI:12973933
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 66)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0072 row: A column: 07
 Seq primer: CGTTGTAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 66.

FEATURES Location/Qualifiers
 1..66
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0072A07"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gil4732114|9b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 19 a 14 c 16 g 17 t
ORIGIN

Query Match 100.0%; Score 9; DB 250; Length 66;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 42 TAATACTAG 34

RESULT 10
AI761414
LOCUS
DEFINITION
w965c06.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2369962 3' similar to TR:075425 075425 ORF3, SPLICEVARIANT_B.
; mRNA sequence.

ACCESSION
VERSION
AI761414
AI761414.1 GI:5177081
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 67)

REFERENCE
AUTHORS
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 1144 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. 67

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2369962"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
18 a 14 c 24 g 11 t

ORIGIN

Query Match 100.0%; Score 9; DB 24; Length 67;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 10 TAATACTAG 18

RESULT 11

AA250291
LOCUS
DEFINITION
mz61g02.r1 Soares mouse lymph node NBMLN Mus musculus cDNA clone
IMAGE:717938 5' mRNA sequence.

ACCESSION
VERSION
AA250291
AA250291.1 GI:1882038
EST.
KEYWORDS
house mouse.
SOURCE
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71)

REFERENCE
AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:443434
Putative full length read
vector to vector length is
Seq primer: -28ml3 rev2 EF from Amersham.
Location/Qualifiers
1. 71

FEATURES
source

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:717938"
/clone_lib="Soares mouse lymph node NBMLN"
/sex="male"
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: p7T3D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'
TGTTACCAATCTGAAGTCGAGCGCGGACTACTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M. Fatima Bonaldo."
32 a 10 c 12 g 17 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 9; DB 4; Length 71;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu Oct 25 13:08:55 2001

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1081639

Seq primer: -40RP from Gibco.

FEATURES
source

Location/Qualifiers
1. .71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3372035"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 32 a 10 c 12 g 17 t
ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
DB 58 TAATACTAG 66

RESULT 14
BE630111/c 71 bp mRNA EST 25-AUG-2000
LOCUS u15f06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:3372035 3', mRNA sequence.

ACCESSION BE630111
VERSION BE630111.1 GI:9912799
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 71)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

REFERENCE
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTS: u15f06.y2
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1081639.

FEATURES
source

Location/Qualifiers
1. .71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3372035"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary

QY 1 taatactag 9
|||||
DB 58 TAATACTAG 66

RESULT 12

AW989354 71 bp mRNA EST 02-JUN-2000
LOCUS u15f06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1512373 5', mRNA sequence.

ACCESSION AW989354

VERSION AW989354

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 71)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:939225

Seq primer: -40RP from Gibco.

FEATURES
source

Location/Qualifiers
1. .71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1512373"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 32 a 10 c 12 g 17 t
ORIGIN

Query Match 100.0%; Score 9; DB 123; Length 71;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
DB 58 TAATACTAG 66

RESULT 13

BE627067

LOCUS u15f06.y2 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:3372035 5', mRNA sequence.

ACCESSION BE627067

VERSION BE627067.1 GI:9907487

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 71)

REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I, and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 17 a 12 c 11 g 31 t
ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
|||||||
Db 13 TAATACTAG 5

RESULT 15
BE688187/c
LOCUS
DEFINITION BE688187 71 bp mRNA EST 11-SEP-2000
IMAGE:3464723 3', mRNA sequence.
ACCESSION BE688187
VERSION
KEYWORDS
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 71)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:1384083.

FEATURES

Location/Qualifiers
1..71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3464723"
/clone_lib="Soares.mammary.gland_NMLMG".
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH108"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I, and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 18 a 12 c 9 g 32 t
ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;
Best Local Similarity 100.0%; Pred. No. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
|||||||
Db 13 TAATACTAG 5

us-09-462-955-4.rst

Thu Oct 25 13:08:55 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:31 ; Search time 180.6 Seconds
(without alignments)
9.434 Million cell updates/sec

Title: US-09-462-955-4

Perfect score: 9
Sequence: 1 taactactag 9

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCPTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	9	100.0	12	2	US-08-418-071-15	Sequence 15, Appl
C 2	9	100.0	30	3	US-09-150-133-33	Sequence 33, Appl
C 3	9	100.0	30	3	US-09-150-141-33	Sequence 33, Appl
C 4	9	100.0	30	4	US-09-374-493-33	Sequence 33, Appl
C 5	9	100.0	30	4	US-09-374-824-33	Sequence 33, Appl
C 6	9	100.0	30	4	US-09-374-492-33	Sequence 33, Appl
C 7	9	100.0	38	3	US-08-817-926-41	Sequence 41, Appl
C 8	9	100.0	41	3	US-08-491-954-106	Sequence 106, Appl
C 9	9	100.0	62	3	US-08-492-076-9	Sequence 9, Appl
C 10	9	100.0	78	2	US-08-680-326-51	Sequence 51, Appl
C 11	9	100.0	101	1	US-07-795-859B-14	Sequence 14, Appl
C 12	9	100.0	101	1	US-08-457-616-14	Sequence 14, Appl
C 13	9	100.0	299	2	US-08-637-759B-31	Sequence 31, Appl
C 14	9	100.0	299	3	US-08-871-355A-31	Sequence 31, Appl
C 15	9	100.0	352	6	5242821-11	Patent No. 5242821
C 16	9	100.0	454	1	US-08-846-134-2	Sequence 2, Appl
C 17	9	100.0	492	1	US-08-470-720-3	Sequence 3, Appl
C 18	9	100.0	573	1	US-08-709-912-18	Sequence 18, Appl
C 19	9	100.0	573	2	US-09-047-370-18	Sequence 18, Appl
C 20	9	100.0	593	3	US-08-705-875A-2	Sequence 2, Appl
C 21	9	100.0	600	5	PCT-US91-02766-24	Sequence 24, Appl
C 22	9	100.0	628	2	US-08-874-186-45	Sequence 45, Appl
C 23	9	100.0	660	3	US-09-284-782-15	Sequence 15, Appl
C 24	9	100.0	663	4	US-08-998-416-187	Sequence 187, Appl
C 25	9	100.0	706	3	US-08-654-025-4	Sequence 4, Appl
C 26	9	100.0	713	1	US-08-234-939-9	Sequence 9, Appl
C 27	9	100.0	713	1	US-08-558-865-9	Sequence 9, Appl

Sequence 849, Appl
Sequence 1138, Appl
Sequence 628, Appl
Sequence 261, Appl
Sequence 11, Appl
Sequence 56, Appl
Sequence 50, Appl
Sequence 289, Appl
Sequence 8, Appl
Sequence 3, Appl
Sequence 47, Appl
Sequence 47, Appl
Sequence 9, Appl
Sequence 48, Appl
Sequence 1, Appl

US-08-418-071-15/c
; Sequence 15, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: other nucleic acid
; ANTI-SENSE: NO
; US-08-418-071-15

ALIGNMENTS

RESULT 1

US-08-418-071-15/c
; Sequence 15, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: other nucleic acid
; ANTI-SENSE: NO
; US-08-418-071-15

Query Match 100.0%; Score 9; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

us-09-462-955-4.rni

Thu Oct 25 13:08:53 2001

```
Qy 1 taatactag 9
    |||||
Db 10 TAATACTAG 2

RESULT 2
US-09-150-133-33/c
; Sequence 33, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-150-133-33

Query Match 100.0%; Score 9; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 3
US-09-150-141-33/c
; Sequence 33, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-150-141-33

Query Match 100.0%; Score 9; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 4
US-09-374-493-33/c
; Sequence 33, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
```

```
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-493-33

Query Match 100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 5
US-09-374-824-33/c
; Sequence 33, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-824-33

Query Match 100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
    |||||
Db 12 TAATACTAG 4

RESULT 6
US-09-374-492-33/c
; Sequence 33, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
```

```

; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 33
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-492-33

```

```

Query Match      100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 taatactag 9
Db 12 TAATACTAG 4

```

```

RESULT 7
US-08-817-926-41
; Sequence 41, Application US/08817926
; Patent No. 6001590
; GENERAL INFORMATION:
; APPLICANT: Kameda, Toshihiro
; APPLICANT: Suda, Hisako
; APPLICANT: Tamai, Yukio
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Kato, No. 6001590no
; APPLICANT: Sakai, Yasuyoshi
; TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINII
; TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,926
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/02597
; FILING DATE: 12-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 234133/1995
; FILING DATE: 12-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 42536/1996
; FILING DATE: 29-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 081356/0112
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
US-08-817-926-41

```

```

Query Match      100.0%; Score 9; DB 3; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 taatactag 9
Db 10 TAATACTAG 18

```

```

RESULT 8
US-08-491-954-106/c
; Sequence 106, Application US/08491954
; Patent No. 6096321
; GENERAL INFORMATION:
; APPLICANT: Girardeau, Jean-Pierre
; APPLICANT: Martin, Christine
; APPLICANT: Mechlin, Marie-Claire
; APPLICANT: Der Vartanian, Maurice
; APPLICANT: Bousquet, Francois
; TITLE OF INVENTION: SUB-UNIT OF CS31A PROTEIN CAPSULE
; TITLE OF INVENTION: MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A
; TITLE OF INVENTION: PROTEIN CAPSULE INCLUDING SUCH A SUB-UNIT, AND
; TITLE OF INVENTION: MICROORGANISMS WHOSE OUTER MEMBRANE CARRIES SUCH
; TITLE OF INVENTION: SUB-UNITS, AND PROCEDURE FOR OBTAINING AND UTILIZING SUCH
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,954
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/01281
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989.6264P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna

```

```
;
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..41
;
; US-08-491-954-106
;
;
; Query Match      100.0%; Score 9; DB 3; Length 41;
; Best Local Similarity 100.0%; Pred. No. 1.4e+03;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 taatactag 9
;      |||||
; Db 9 TAATACTAG 1
;
;
; RESULT 9
; US-08-492-076-9/c
; Sequence 9, Application US/08492076A
; Patent No. 6060064
; GENERAL INFORMATION:
; APPLICANT: Adams, Sally E.
; APPLICANT: Burus, Nigel R.
; APPLICANT: Richardson, Simon M.
; TITLE OF INVENTION: No. 6060064el Proteinaceous Particles
; FILE REFERENCE: 10180.60968
; CURRENT APPLICATION NUMBER: US/08/492.076A
; CURRENT FILING DATE: 1995-06-28
; EARLIER APPLICATION NUMBER: PCT/GB93/02656
; EARLIER FILING DATE: 1993-12-24
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 62
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Partial V3
; OTHER INFORMATION: loop of HIV-1 further comprising restriction
; OTHER INFORMATION: enzyme sites.
;
; US-08-492-076-9
;
; Query Match      100.0%; Score 9; DB 3; Length 62;
; Best Local Similarity 100.0%; Pred. No. 1.3e+03;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 taatactag 9
;      |||||
; Db 9 TAATACTAG 1
;
;
; RESULT 10
; US-08-680-326-51/c
; Sequence 51, Application US/08680326
; Patent No. 5925733
; GENERAL INFORMATION:
; APPLICANT: ROSE, TIMOTHY M.
; APPLICANT: BOSCH, MARIX
; APPLICANT: STRAND, KURT
; APPLICANT: TODARO, GEORGE J.
; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
; TITLE OF INVENTION: FIBROMATOSIS
; NUMBER OF SEQUENCES: 152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
;
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680.326
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Schliff, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-680-326-51
;
; Query Match      100.0%; Score 9; DB 2; Length 78;
; Best Local Similarity 100.0%; Pred. No. 1.2e+03;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 taatactag 9
;      |||||
; Db 37 TAATACTAG 29
;
;
; RESULT 11
; US-07-795-859B-14
; Sequence 14, Application US/07795859B
; Patent No. 5422262
; GENERAL INFORMATION:
; APPLICANT: Anderson, Stefan
; APPLICANT: Russell, David W.
; TITLE OF INVENTION: Steroid 5 $\alpha$ -Reductases
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White and Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/795.859B
; FILING DATE: 18-NOV-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:260/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7677
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 101 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; US-07-795-859B-14
```

Query Match 100.0%; Score 9; DB 1; Length 101;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
 |||||
 Db 18 TAATACTAG 26

RESULT 12

US-08-457-616-14
 ; Sequence 14, Application US/08457616
 ; Patent No. 5679521
 ; GENERAL INFORMATION:
 ; APPLICANT: Anderson, Stefan
 ; APPLICANT: Russell, David W.
 ; TITLE OF INVENTION: Steroid 5 α -Reductases
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White and Durkee
 ; STREET: P O Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII-DOS
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/457,616
 ; FILING DATE: 01-JUN-1995
 ; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/795,859
 ; FILING DATE: 18-NOV-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: UTSD:260/PAR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 320-7200
 ; TELEFAX: (512) 474-7677

INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 101 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-457-616-14

Query Match 100.0%; Score 9; DB 1; Length 101;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
 |||||
 Db 18 TAATACTAG 26

RESULT 13

US-08-637-759B-31
 ; Sequence 31, Application US/08637759B
 ; Patent No. 5876931
 ; GENERAL INFORMATION:
 ; APPLICANT: David William Holden
 ; TITLE OF INVENTION: Identification of Genes
 ; NUMBER OF SEQUENCES: 501

CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Patrea L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; STREET: 1201 West Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/637,759B
 ; FILING DATE: 03-MAY-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/GB95/02875
 ; FILING DATE: 11-DEC-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pabst, Patrea L.
 ; REGISTRATION NUMBER: 31,284
 ; REFERENCE/DOCKET NUMBER: RPMS 101
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (404) 873-8794
 ; TELEFAX: (404) 873-8795
 ; INFORMATION FOR SEQ ID NO: 31:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 299 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Partial sequence of Salmonella typhimurium
 ; ORGANISM: virulence gene
 ; US-08-637-759B-31

Query Match 100.0%; Score 9; DB 2; Length 299;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 taatactag 9
 |||||
 Db 44 TAATACTAG 52

RESULT 14

US-08-871-355A-31
 ; Sequence 31, Application US/08871355A
 ; Patent No. 6015669
 ; GENERAL INFORMATION:
 ; APPLICANT: David William Holden
 ; TITLE OF INVENTION: Identification of Genes
 ; NUMBER OF SEQUENCES: 501
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Patrea L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; STREET: 1201 West Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

Thu Oct 25 13:08:53 2001

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871.355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 299 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Partial sequence of Salmonella typhimurium
ORGANISM: virulence gene
US-08-871-355A-31

Query Match 100.0%; Score 9; DB 3; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 44 TAATACTAG 52

RESULT 15
5242821-11/c
Patent No. 5242821
APPLICANT: PALVA, LIKKA; SIBAKOV, MERV
TITLE OF INVENTION: LACTOCOCCUS PROMOTER AND SIGNAL
SEQUENCES FOR EXPRESSION IN BACTERIA
NUMBER OF SEQUENCES: 27
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/377,450
FILING DATE: 10-JUL-1989
SEQ ID NO: 11
LENGTH: 352
5242821-11

Query Match 100.0%; Score 9; DB 6; Length 352;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 taatactag 9
|||||
Db 139 TAATACTAG 131

Search completed: October 24, 2001, 10:00:31
Job time: 216 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:28 ; Search time 393.52 seconds
(without alignments)
14.360 Million cell updates/sec

Title: US-09-462-955-5

Perfect score: 9

Sequence: 1 ctatgatta 9

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	100.0	18	19	AAV33149
2	9	100.0	18	19	AAV33146
3	9	100.0	22	21	AA30640
4	9	100.0	22	21	AA250574
5	9	100.0	22	22	AA31603
6	9	100.0	23	19	AAV45715
7	9	100.0	28	21	AA261141
8	9	100.0	30	20	AA387693
9	9	100.0	30	21	AA299717
10	9	100.0	31	15	AAQ56057
11	9	100.0	31	20	AA302688
					Plasmodium falciparum
					Plasmodium vivax
					Human G protein-co
					3' PCR primer-1 to
					Double-stranded DN
					Human NKCC2 gene e
					Reverse PCR primer
					Human TST-1 cDNA
					PCR primer for cDN
					Coconut Foliar Dec
					Gemini virus DNA f

C	12	9	100.0	31	20	AA302686	Gemini virus DNA f
	13	9	100.0	32	18	AA30052	Primer for chicken
	14	9	100.0	35	18	AA30052	Pisum sativum plas
	15	9	100.0	35	21	AA30052	Z. paucivorans 16S
	16	9	100.0	36	20	AA30052	PCR primer PC641 f
	17	9	100.0	36	18	AA30052	PCR primer PF668 f
	18	9	100.0	38	20	AA30052	PCR primer PF668 f
	19	9	100.0	38	21	AA30052	Primer PCR2AS for
	20	9	100.0	38	22	AA30052	Primer PCR2AS for
	21	9	100.0	38	22	AA30052	CmR gene/PorA prom
	22	9	100.0	38	22	AA30052	CmR gene/PorA prom
	23	9	100.0	39	21	AA30052	Human G protein-co
	24	9	100.0	39	21	AA30052	3' PCR primer-2 to
	25	9	100.0	41	15	AA30052	CmR gene/PorA prom
	26	9	100.0	41	15	AA30052	Linker coding for
	27	9	100.0	53	21	AA30052	T. reesei xylanase
	28	9	100.0	62	15	AA30052	Oligo encoding N-t
	29	9	100.0	67	21	AA30052	T. reesei xylanase
	30	9	100.0	74	22	AA30052	Variable oligonucle
	31	9	100.0	93	16	AA30052	Human gene signatu
	32	9	100.0	98	21	AA30052	Human secreted pro
	33	9	100.0	101	16	AA30052	Human type 1 stereo
	34	9	100.0	110	18	AA30052	Staphylococcus aur
	35	9	100.0	112	21	AA30052	Human secreted pro
	36	9	100.0	114	21	AA30052	Soybean microsatel
	37	9	100.0	116	21	AA30052	Human secreted pro
	38	9	100.0	127	18	AA30052	Staphylococcus aur
	39	9	100.0	138	21	AA30052	Human secreted pro
	40	9	100.0	149	21	AA30052	Human secreted pro
	41	9	100.0	151	21	AA30052	Human secreted exp
	42	9	100.0	152	16	AA30052	Human secreted pro
	43	9	100.0	152	16	AA30052	Human gene signatu
	44	9	100.0	159	21	AA30052	Human secreted exp
	45	9	100.0	165	21	AA30052	Human secreted pro
				177	19	AA30052	Human biallelic po

ALIGNMENTS

RESULT	1
AAV33149	ID AAV33149 standard; DNA; 18 BP.
XX	AC AAV33149;
XX	AC AAV33149;
DT	07-DEC-1998 (first entry)
XX	Plasmodium falciparum coxi gene PCR primer PFL.
DE	Malaria; infection; therapy; diagnosis; mitochondrion; coxi gene;
XX	cytochrome C oxidase; PCR; primer; ss.
KW	Synthetic.
XX	Plasmodium falciparum.
OS	WO9835057-A1.
XX	13-AUG-1998.
PD	05-FEB-1998; 98WO-IB00212.
XX	26-SEP-1997; 97AU-0009481.
PR	06-FEB-1997; 97AU-0004953.
PR	21-APR-1997; 97AU-0006329.
XX	(MOLE-) INST MOLECULAR & CELL BIOLOGY.
PA	(OYSL-) UNIV SINGAPORE NAT.
XX	Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;
PI	WPI; 1998-447251/38.
XX	Detecting Plasmodium infection from hybridisation with
XX	
PT	

PT extrachromosomal element - providing genus or species specific
 PT diagnosis with few false negatives, in humans or animals
 XX
 XX Claim 6; Page 66; 120pp; English.

XX This synthetic oligonucleotide sequence (PFI) is derived from the
 CC Plasmodium falciparum mitochondrial cytochrome C oxidase (coxI)
 CC gene. In a claimed method, Plasmodium is detected in a human or
 CC animal sample (especially blood or dried blood) by treating it, or
 CC derived nucleic acid, with a Plasmodium extrachromosomal genetic
 CC element or derived nucleic acid (A) and detecting any hybridisation.
 CC (A) can include the PSI-PL470, PLH-PPH, PRB or PWQ gene (see
 CC AAV33135-38), the coxI gene or nucleic acids derived from them. Also
 CC new are (A)-specific probes and primers (see AAV33139-56). The method
 CC is used to diagnose Plasmodium infection. The high degree of
 CC similarity between (A) from different species allows development of
 CC genus- or species-specific assays that result in fewer false
 CC negatives than known methods (typically 1% against 3%). Primer PFI
 CC can be used with another primer (see AAV33145) for the species-
 CC specific detection of P. falciparum.

XX Sequence 18 BP; 4 A; 3 C; 3 G; 8 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
 Db 2 ctagtattta 10

RESULT 2

AAV33146
 ID AAV33146 standard; DNA; 18 BP.

XX AAV33146;

XX 07-DEC-1998 (first entry)

Plasmodium vivax coxI gene PCR primer PV1.

XX Malaria; infection; therapy; diagnosis; mitochondrion; coxI gene;
 KW cytochrome C oxidase; PCR; primer; ss.

XX Synthetic.

OS Plasmodium vivax.

XX WO9835057-A1.

XX 13-AUG-1998.

XX 05-FEB-1998; 98WO-IB00212.

XX 26-SEP-1997; 97AU-0009481.

XX 06-FEB-1997; 97AU-0004953.

XX 21-APR-1997; 97AU-0006329.

XX (MOLE-) INST MOLECULAR & CELL BIOLOGY.

XX (UYSI-) UNIV SINGAPORE NAT.

XX Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;

XX WPI; 1998-447251/38.

XX Detecting Plasmodium infection from hybridisation with
 PT extrachromosomal element - providing genus or species specific
 PT diagnosis with few false negatives, in humans or animals

XX Claim 6; Page 65; 120pp; English.

XX This synthetic oligonucleotide sequence (PVI) is derived from the

CC Plasmodium vivax mitochondrial cytochrome C oxidase (coxI) gene.
 CC In a claimed method, Plasmodium is detected in a human or animal
 CC sample by treating it, or derived nucleic acid, with a Plasmodium
 CC extrachromosomal genetic element or derived nucleic acid (A) and
 CC detecting any hybridisation. (A) can include the PSI-PL470,
 CC PLH-PPH, PRB or PWQ gene (see AAV33135-38), the coxI gene or nucleic
 CC acids derived from them. Also new are (A)-specific probes and
 CC primers (see AAV33139-56). The method is used to diagnose Plasmodium
 CC infection. The sample can be blood or dried blood. The high
 CC degree of similarity between (A) from different species allows
 CC development of genus- or species-specific assays that result in
 CC fewer false negatives than known methods (typically 1% against 3%).
 CC Primer PVI can be used with another primer (see AAV33145) for the
 CC species-specific detection of P. vivax.

XX Sequence 18 BP; 3 A; 3 C; 3 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 19; Length 18;
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
 Db 2 ctagtattta 10

RESULT 3

AAA30640/C
 ID AAA30640 standard; DNA; 22 BP.

XX AAA30640;

XX 21-AUG-2000 (first entry)

XX Human G protein-coupled receptor GHSR PCR primer, SEQ ID NO:84.

XX G protein-coupled receptor; GPCR; constitutively active;
 KW intracellular loop 3; transmembrane domain 6; drug screening;
 KW agonist; antagonist; PCR primer; ss.

XX Homo sapiens.

OS WO200022129-A1.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-US23938.

XX 13-OCT-1998; 98US-0170496.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw CW;

XX WPI; 2000-329165/28.

XX Non-endogenous constitutively activated human G protein-coupled
 PT receptors, useful for identifying agonists for use as pharmaceutical
 PT agents

XX Example 1; Page 42; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions
 CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAV90643-
 CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAV30709-A30743
 CC and AAV30775-A30779). The mutant proteins of the invention contain a
 CC mutation in a portion of the protein comprising intracellular loop 3
 CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
 CC is substituted for an endogenous residue in IC3 at a position 16 amino
 CC acids N-terminal of an endogenous proline in TM6 to form a sequence
 CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
 CC or Ala, and is preferably Lys. When the endogenous residue at this

CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
 CC The 15 amino acid stretch between the substituted amino acid and the Pro
 CC may be endogenous, non-endogenous, or a mixture of endogenous and
 CC non-endogenous residues. The constitutively active GPCRs are useful for
 CC identifying antagonists, agonists and partial agonists for use as
 CC pharmaceutical agents. The mutant proteins are also useful in research
 CC settings for elucidating the roles of the receptors in normal and
 CC diseased conditions. Antagonists for a particular GPCR are useful for
 CC treating diseases and disorders associated with that receptor. Because
 CC the novel mutant GPCRs are constitutively active, they can be used
 CC directly for screening of compounds without the need for endogenous
 CC ligands. The present sequence represents a PCR primer used in an
 CC exemplification of the invention to isolate cDNA encoding a human
 CC wild-type GPCR for cloning, an exemplification of the invention. The GPCR
 CC cDNA was subjected to site-directed mutagenesis (SDM) to generate DNA
 CC encoding the corresponding mutant of the invention.

XX Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9
 Db 17 CTAGTATTATA 9

RESULT 4

AAZ50574/C
 ID AAZ50574 standard; DNA; 22 BP.

XX AAZ50574;

XX 20-JUN-2000 (first entry)

XX 3' PCR primer-1 to amplify human GHSR DNA.

XX G protein-coupled orphan receptor; GPCR; agonist; G protein;
 KW GPCR fusion protein; inverse agonist; drug; treatment; PCR primer;
 KW GHSR; G protein-coupled receptor; human; ss.

XX Homo sapiens.

XX WO200006597-A2.

XX 10-FEB-2000.

XX 30-JUL-1999; 99WO-US17425.

XX 31-JUL-1998; 98US-0094879.

XX 30-OCT-1998; 98US-0106300.

XX 04-DEC-1998; 98US-0110906.

XX 26-FEB-1999; 99US-0121851.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw C, Lin I, Lowitz K, Chen R;

XX WPI; 2000-195260/17.

XX Identification of a compound useful as a therapeutic agent, comprises
 PT identifying a compound against constitutively activated G
 PT protein-coupled orphan receptors -

XX Example 2; Page 27; 123pp; English.

XX The patent discloses a method of identifying agonists and inverse or
 CC partial agonists to the endogenous, constitutively activated
 CC G protein-coupled orphan receptors (GPCRs), by contacting them with a
 CC GPCR fusion protein comprising a GPCR and a G protein. Determining
 CC expression of GPCRs in tissue samples can be used to identify related

CC diseases. Inverse agonists to these receptors can be used as drugs for
 CC treating GPCR-related diseases. The present sequence is a 3' PCR primer,
 CC used in primary PCR reaction to amplify human GHSR DNA from
 CC hippocampus cDNA template.

XX Sequence 22 BP; 7 A; 3 C; 2 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 22;
 Best Local Similarity 100.0%; Pred. No. 3.8e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9
 Db 17 CTAGTATTATA 9

RESULT 5

AAF31603/C

ID AAF31603 standard; DNA; 22 BP.

XX AAF31603;

XX 09-APR-2001 (first entry)

XX Double-stranded DNA tag PCR primer phag2.

XX attenuated microorganism; signature tagged transposon mutant;
 KW mutant library; mycobacterial infection; actinomycetales;
 KW antibacterial; immunostimulant; vaccine; PCR primer; ss.

XX Unidentified.

XX WO200102555-A1.

XX 11-JAN-2001.

XX 06-JUL-2000; 2000WO-IB00950.

XX 06-JUL-1999; 99US-0142982.

XX 08-JUL-1999; 99US-0142833.

XX (INSP) INST PASTEUR.

XX Gicquel B, Guilhot C, Camacho L;

XX WPI; 2001-091804/10.

XX Screening a mutant library for mutants unable to grow under specific
 PT conditions and for identifying loci involved in pathogenicity,
 PT comprises using signature tagged transposon mutagenesis -

XX Example 4; Page 19; 159pp; English.

XX The present sequence is given in a specification relating to a method for
 CC screening a library of mutants. The method comprises constructing a
 CC library with insertions in genes and/or regulatory regions of the
 CC organisms of interest, where the insertion contains a tag and/or a
 CC transposon associated with a tag. The mutants are identified by
 CC hybridisation of the tags to known sequences. The method is useful for
 CC treating an individual suffering from a mycobacterial infection,
 CC suspected of being infected with a Mycobacterium, or having been
 CC exposed to an infectious Mycobacterium. It is also useful for
 CC identifying and isolating mutants of Mycobacterium. It is also useful for
 CC compounds that have antibiotic activity. The method is used to identify
 CC mutants of microorganisms, preferably an actinomycetales, such as
 CC M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
 CC M. paratuberculosis, that is unable to grow under specific conditions.
 CC It is especially useful for identifying loci involved in pathogenicity.
 CC It is useful in constructing vaccines. The method can be used to screen
 CC multiple libraries concurrently. It can screen libraries of different
 CC organisms or different strains of the same organism. The present
 CC sequence was used to create the tag used in the construction of tagged

CC mutants.

XX Sequence 22 BP; 7 A; 5 C; 1 G; 9 T; 0 other;

SQ Query Match 100.0%; Score 9; DB 22; Length 22;

Best Local Similarity 100.0%; Pred. No. 3.8e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

|||||

Db 11 CTAGTATTA 3

RESULT 6

AAV45715/C

ID AAV45715 standard; DNA; 23 BP.

XX AC AAV45715;

XX 21-DEC-1998 (first entry)

XX Human NKCC2 gene exon 19 forward primer hNKCC2ex19.

XX Na-K-2Cl cotransporter; NKCC2; human; Bartter's syndrome;

XX ion transport; hypokalaemic alkalosis; hypercalciuria;

XX nephrocalcinosis; diagnosis; therapy; SSCP; primer; ss.

XX Synthetic.

XX Homo sapiens.

XX WO9829431-A1.

XX 09-JUL-1998.

XX 19-DEC-1997; 9TWO-US23553.

XX 31-DEC-1996; 96US-0778052.

XX (UYVA) UNIV YALE.

XX Lifton RP, Simon DB;

XX WPI; 1998-388029/33.

XX Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2Cl

XX cotransporter genes - useful for developing products for the

XX diagnosis and treatment of ion transport disorders, e.g. Gitelman's

XX Syndrome or Bartter's Syndrome

XX Example 2; Page 65; 105pp; English.

XX Primers hNKCC2ex19 forward and reverse (see AAV45715 and AAV45716,

XX respectively) are designed to amplify exon 19 of the human NKCC2

XX gene (see AAV40562) that codes for Na-K-2Cl cotransporter NKCC2

XX protein (see AAW29683). Both primers are located within introns of

XX the gene. 27 Sets of specific primers (see AAV45677-V45730) were

XX used for SSCP analysis of NKCC2. Amplified products were analysed

XX for molecular variants by electrophoresis, and identified variants

XX were sequenced. Complete linkage of Bartter's syndrome with NKCC2

XX was demonstrated. Identification of the molecular basis of

XX Bartter's syndrome allows for the genetic diagnosis of this

XX disorder. The invention provides products and methods useful for

XX diagnosis and treatment of Bartter's syndrome and other ion

XX transport disorders.

XX Sequence 23 BP; 8 A; 4 C; 5 G; 6 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 9; DB 19; Length 23;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

|||||

Db 10 CTAGTATTA 2

RESULT 7

AAZ61141

ID AAZ61141 standard; DNA; 28 BP.

XX AC AAZ61141;

XX 30-MAY-2000 (first entry)

XX Reverse PCR primer used to amplify DNA encoding the Shiga toxin.

XX Chemokine receptor; ligand; inflammatory response; immune effector cell;

XX secondary tissue damage; central nervous system injury; Shiga;

XX CNS inflammatory disease; neurodegenerative disorder; heart disease;

XX inflammatory eye disease; inflammatory bowel disease; PCR primer;

XX inflammatory joint disease; inflammatory kidney; renal disease;

XX inflammatory lung disease; inflammatory nasal disease; thyroiditis;

XX inflammatory thyroid disease; cytokine-regulated cancer; ss.

XX Shigella dysenteriae.

XX WO200004926-A2.

XX 03-FEB-2000.

XX 21-JUL-1999; 99WO-CA00659.

XX 22-JUL-1998; 98US-0120523.

XX (OSPR-) OSPREY PHARM LTD.

XX McDonald JR, Coggins PJ;

XX WPI; 2000-182542/16.

XX A new therapeutic agent comprising a conjugate for treating secondary

XX tissue damage and other disease conditions like Alzheimer's disease,

XX stroke, Parkinson's disease and atherosclerosis

XX Example 1; Page 135; 204pp; English.

XX PCR primers AAZ61140-42 were used to amplify nucleic acids encoding

XX the Shiga toxin. The toxin can be incorporated into the conjugates

XX of the invention. The specification describes a conjugate, comprising

XX a targeted agent and a chemokine receptor ligand. The conjugate

XX binds to a chemokine receptor resulting in internalisation of the

XX targeted agent in cells bearing the receptor. The conjugates are

XX used for formulating a medicament or for treating disorders associated

XX with inflammatory responses resulting from activation, proliferation

XX and migration of immune effector cells. The disorders or disease states

XX comprise secondary tissue damage such as central nervous system (CNS)

XX injury, CNS inflammatory diseases, neurodegenerative disorders, heart

XX disease, inflammatory eye diseases, inflammatory bowel diseases,

XX inflammatory joint diseases, inflammatory kidney or renal diseases,

XX inflammatory lung diseases, inflammatory nasal diseases, inflammatory

XX thyroid disease such as thyroiditis, or cytokine-regulated cancers.

Query Match 100.0%; Score 9; DB 21; Length 28;

Best Local Similarity 100.0%; Pred. No. 3.7e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

|||||

Db 11 ctagtattta 19

```

RESULT 8
AAZ87693
ID AAX87693 standard; DNA; 30 BP.
XX
AC AAX87693;
XX
DT 26-OCT-1999 (first entry)
XX
DE Human TPST-1 cDNA bottom strand PCR primer.
XX
KW TPST-1; tyrosylprotein sulfotransferase; tyrosine O-sulfation;
KW post-translation modification; human; PCR; primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9938980-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-0501756.
XX
PR 09-SEP-1998; 98US-0150141.
PR 29-JAN-1998; 98US-0072994.
PR 09-SEP-1998; 98US-0150133.
XX
PA (OKLA ) UNIV OKLAHOMA STATE.
XX
PI Moore KL;
XX
DR WPI; 1999-494094/41.
XX
PT Human, murine and Caenorhabditis elegans tyrosylprotein
PT sulfotransferases, useful for post-translational tyrosine sulfation
XX
PS Example; Page 39; 123pp; English.
XX
CC This bottom strand primer was used with a top strand primer (see
CC AAX87692) in the PCR amplification of the human tyrosylprotein
CC sulfotransferase (TPST-1) coding sequence (see AAX87686) using ESR
CC clone #116978 as template. The primers introduce a 5' BamHI site
CC and a 3' SpeI site. The amplified cDNA was ligated into vector
CC pCDNA3.1(+), and TPST-1 (see AAY06623) was expressed as a fusion
CC protein containing an N-terminal epitope for HPC4 (see AAY06630)
CC in mammalian cells. TPST-1 is useful for the post-translational
CC tyrosine O-sulfation of proteins and peptides.
XX
SQ Sequence 30 BP; 5 A; 9 C; 6 G; 10 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9
Db 4 ctagtattta 12

RESULT 9
AAZ99717
ID AAX99717 standard; DNA; 30 BP.
XX
AC AAX99717;
XX
DT 12-JUL-2000 (first entry)
XX
DE PCR primer for cDNA encoding human tyrosylprotein sulfotransferase 1.
XX
KW Human; tyrosylprotein sulfotransferase 1; TPST-1; tyrosine O-sulfation;
KW P-selectin glycoprotein ligand-1; PSGL-1; anti-inflammatory;
KW neutrophil binding; PCR primer; ss.
XX
XX

Query Match 100.0%; Score 9; DB 21; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9
Db 4 ctagtattta 12

RESULT 10
AAQ56057/c
ID AAQ56057 standard; DNA; 31 BP.
XX
AC AAQ56057;
XX
DT 12-AUG-1994 (first entry)
XX
DE Coconut Foliar Decay Virus promoter stem-loop.
XX
KW Coconut Foliar Decay Virus; CFDV; strong promoter; tissue-specific;
KW phloem-specific; stem-loop structure; transgenic plant; ds.
XX
OS Coconut Foliar Decay Virus.
XX
PH Key Location/Qualifiers
FT stem_loop 1..31
FT /*tag= a
FT /function= promoter
FT /note= "loop has homology to geminivirus sequence"
XX
XX DE4306832-C.
XX
PD 24-FEB-1994.

```

XX PF 04-MAR-1993; 93DE-4306832.
XX PR 04-MAR-1993; 93DE-4306832.
XX PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Becker D, Randles JW, Rohde W, Salamini F;
XX DR WPI; 1994-058406/08.
XX PT Use of coconut foliar decay virus DNA as promoter - for
XX PT tissue-specific gene expression in transgenic plants
XX PS Claim 1; Fig 2; 8pp; German.
XX CC A DNA fragment from the CFV genome can be used as a phloem-specific
XX CC promoter in the construction of transgenic plants. The promoter is
XX CC strong; it has 30-50% of the activity of the CaMV 35S promoter in
XX CC tobacco plants.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 15; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6e+03; Indels 0; Caps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 ctagtattata 9
DB 20 CTAGTATTATA 12

RESULT 11
AAX02688/C
ID AAX02688 standard; DNA; 31 BP.
XX AC AAX02688;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX OS Gemini virus.
XX FH Location/Qualifiers
FT Key 1..31
FT stem_loop /*tag= a
FT misc_binding 1..11
FT /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31
FT /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX DE19730502-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030502.
XX PF 16-JUL-1997; 97DE-1030502.
XX PR (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096863/09.
XX DR Coconut foliar decay virus promoters - for gene expression in
XX PT

PT bacteria and yeasts
XX Disclosure; Fig 2; 14pp; German.
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFV DNA fragment is useful as a bacterial or yeast promoter, as a
XX CC promoter for tissue-specific (especially phloem-specific) gene expression
XX CC in plants and for production of chimeric constructs for transient or
XX CC stable expression. Certain fragments of CFV DNA have stronger promoter
XX CC activity in E. coli than the CaMV 35S promoter.
XX SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.6e+03; Indels 0; Caps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 ctagtattata 9
DB 20 CTAGTATTATA 12

RESULT 12
AAX02686/C
ID AAX02686 standard; DNA; 31 BP.
XX AC AAX02686;
XX DT 10-MAY-1999 (first entry)
XX DE Gemini virus DNA fragment stem loop.
XX KW Stem loop; coconut foliar decay virus; CFV; bacterial promoter;
XX KW yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.
XX OS Gemini virus.
XX FH Location/Qualifiers
FT Key 1..31
FT stem_loop /*tag= a
FT misc_binding 1..11
FT /*tag= b
FT /*note= "Region binds to nucleotides 21 to 31"
FT misc_binding 21..31
FT /*tag= c
FT /*note= "Region binds to nucleotides 1 to 11"
XX DE19730535-A1.
XX PN 21-JAN-1999.
XX PD 16-JUL-1997; 97DE-1030535.
XX PF 16-JUL-1997; 97DE-1030535.
XX PR (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PA Becker D, Hehn A, Randles JW, Rohde W, Salamini F;
XX PI WPI; 1999-096867/09.
XX DR Coconut foliar decay virus promoters - for gene expression in
XX PT plants, bacteria and yeasts
XX PS Disclosure; Fig 2; 12pp; German.
XX CC This invention describes a coconut foliar decay virus (CFDV) DNA fragment
XX CC that includes the stem-loop structure of CFV DNA but lacks the
XX CC translation start codons of open reading frames ORF1 and/or ORF2. The
XX CC new CFV DNA fragment is useful as a bacterial or yeast promoter, as a

CC promoter for tissue-specific (especially phloem-specific) gene expression
 CC in plants and for production of chimeric constructs for transient or
 CC stable expression. Certain fragments of CPDV DNA have stronger promoter
 CC activity in *E. coli* than the CaMV 35S promoter.
 XX
 SQ Sequence 31 BP; 5 A; 11 C; 11 G; 4 T; 0 other;

Query Match 100.0%; Score 9; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
 |||||
 Db 20 CTAGTATTATA 12

RESULT 13

AAT90052
 ID AAT90052 standard; DNA; 32 BP.

XX
 AC AAT90052;

XX
 DT 27-MAR-1998 (first entry)

XX
 DE Primer for chicken urokinase.cDNA.

XX
 KW Low molecular weight; chicken; urokinase; reporter; screening;
 PCR primer; ss.

XX
 OS Gallus domesticus.

XX
 PN W09734144-AL.

XX
 PD 18-SEP-1997.

XX
 PF 11-MAR-1997; 97WO-EP01219.

XX
 PR 20-DEC-1996; 96DE-4054616.

XX
 PR 14-MAR-1996; 96DE-4011780.

XX
 PA (SCHD) SCHERING AG.

XX
 PI Langer G, Schleuning W., Toschi L;

XX
 DR WPI; 1997-470977/43.

XX
 PT Determining effect of test substance on cultured cells - by
 PT transfecting cells with plasmid having gene for low molecular weight
 PT hen urokinase as reporter, which can be detected by plasminogen
 PT activation and colour forming or fluorochrome substrate reaction
 XX

PS Example 1; Page 16; 32pp; German.

XX
 CC The present sequence was used in the development of a method for
 CC determining the effect of test substance in a cell culture system.
 CC The method comprises preparing a plasmid having an expression
 CC cassette including the low molecular weight gene for chicken
 CC urokinase as a reporter, transfecting eukaryotic cells
 CC with the plasmid, inducing the cells with the test substance,
 CC harvesting the cell supernatant and detecting urokinase by adding
 CC chicken plasminogen (this addition precisely defines the start of
 CC the reaction), so that this is converted to plasmin by the
 CC urokinase, and detecting the plasmin in the supernatant with a
 CC highly sensitive chromogenic or fluorochromic substrate. The method
 CC can be used to screen new compounds, non-steroidal glucocorticoids
 CC or anti-glucocorticoids, determine receptors present in cells,
 CC particularly steroid hormone receptors and investigate promoter
 CC dependent modulated transcription processes. The method does not
 CC require radioisotopes or extraction of reporter gene products, is
 CC sensitive, non-destructive and can be performed in microtitre
 CC plates. Urokinase is small (so easily separated from endogenous
 CC proteases by gel electrophoresis), and unlike other reporter gene

CC products is stable in culture supernatant, permitting direct
 CC measurement. The method can be calibrated using a precise number of
 CC transfected cells, avoiding difficulties of calibration based on
 CC enzymatic activity. The urokinase/plasminogen reaction represents a
 CC long lived amplification system, which is well characterised.
 XX

SQ Sequence 32 BP; 5 A; 7 C; 11 G; 9 T; 0 other;

Query Match 100.0%; Score 9; DB 18; Length 32;
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
 |||||
 Db 12 ctagtattata 20

RESULT 14

AAT77100

ID AAT77100 standard; DNA; 35 BP.

XX
 AC AAT77100;

XX
 DT 20-JAN-1998 (first entry)

XX
 DE Pisum sativum plastocyanin promoter sub-sequence.

XX
 KW plastocyanin; promoter; enhancer; reporter gene; pea;
 transgenic plant; ds.

XX
 OS Pisum sativum.

XX
 FH Key Location/Qualifiers
 FT enhancer 1..31

FT
 FT misc_feature 1..4

FT
 FT /tag= a

FT
 FT /tag= b

FT
 FT /label= sticky_end

FT
 FT /note= "The 5' end of complementary strand
 overhangs the 3' end of this strand by the
 sequence 5'-GATC-3'."

XX
 PN W09720056-A2.

XX
 PD 05-JUN-1997.

XX
 PF 26-NOV-1996; 96WO-GB02910.

XX
 PR 29-NOV-1995; 95GB-0024350.

XX
 PA (ADTE-) ADVANCED TECHNOLOGIES CAMBRIDGE.

XX
 PI Gray JC, Sandhu JS, Webster CI;

XX
 DR WPI; 1997-310610/28.

XX
 PT Enhancer from pea plastocyanin promoter region - used to produce
 PT transgenic plants having increased expression of a desired gene

PS Claim 11; Figure 7; 46pp; English.

XX
 CC This sequence represents the pea plastocyanin promoter enhancer
 CC which is a sub-sequence of AAT77099. This enhancer has been used to
 CC increase the expression of the GUS reporter gene in tobacco where the
 CC GUS gene was under the control of the PetE promoter and 35S CaMV
 CC promoter. The enhancer was also used joined to the -330 to +1 region of
 CC the patatin promoter PS20 which was joined to the GUS reporter and
 CC transformed into potato using A.tumefaciens. Expression of the gene
 CC promoter of one or more genes of a plant increases in green or non-green
 CC tissues, especially roots, tubers, seeds, stems, flowers or leaves and
 CC transformed plants may be dicotyledonous or monocotyledonous. This
 CC enhancer may operate both in normal and reverse directions and can be

CC attached to the promoter or terminator of the gene to be expressed.
XX
SQ Sequence 35 BP; 17 A; 4 C; 2 G; 12 T; 0 other;

Query Match 100.0%; Score 9; DB 18; Length 35;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
| | | | | | | |
Db 11 ctagtattta 19

RESULT 15
AAA39333
ID AAA39333 standard; DNA; 35 BP.
XX
AC AAA39333;
XX 11-SEP-2000 (first entry)
XX Z. paucivorans 16S rRNA and 23S rRNA gene spacer region SEQ ID NO:8.
XX 16S rRNA; 23S rRNA; spacer region; detection; microbe; beer;
KW brewery; ds.
XX Zymophilus paucivorans.
XX
XX JP2000106881-A.
XX 18-APR-2000.
XX
XX 08-OCT-1998; 98JP-0286697.
XX 08-OCT-1998; 98JP-0286697.
XX (ASAK) ASAHI BREWERIES LTD.
XX WPI; 2000-353477/31.
XX
XX A gene sequence spacer region between a gene encoding 16S rRNA and a
PT gene encoding 23S rRNA, useful for detection of microbes -
XX
XX Claim 8; Page 2; 18pp; Japanese.
XX
XX The present invention describes gene sequences of spacer region between
CC a gene encoding 16S rRNA and a gene encoding 23S rRNA for the detection
CC of microbes. Also described is a method for the detection of microbes
CC in which an oligonucleotide prepared from a spacer region as described
CC above is made to function as a primer for nucleic acid synthesis and
CC treated by gene amplification. The method is used for the detection of
CC microbes found in beer breweries. The present sequence represents a
CC specifically claimed Zymophilus paucivorans 16S rRNA and 23S rRNA spacer
CC region DNA sequence, from the present invention.
XX
SQ Sequence 35 BP; 11 A; 3 C; 7 G; 14 T; 0 other;

Query Match 100.0%; Score 9; DB 21; Length 35;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9
| | | | | | | |
Db 7 ctagtattta 15

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:31 ; Search time 180.6 seconds
(without alignments)
9.434 Million cell updates/sec

Title: US-09-462-955-5
Perfect score: 9
Sequence: 1 ctgattatta 9
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	9	100.0	12	2	US-08-418-071-15		Sequence 15, Appl
2	9	100.0	30	3	US-09-150-133-33		Sequence 33, Appl
3	9	100.0	30	3	US-09-150-141-33		Sequence 33, Appl
4	9	100.0	30	4	US-09-374-493-33		Sequence 33, Appl
5	9	100.0	30	4	US-09-374-824-33		Sequence 33, Appl
6	9	100.0	30	4	US-09-374-492-33		Sequence 33, Appl
7	9	100.0	38	3	US-08-817-926-41		Sequence 41, Appl
8	9	100.0	41	3	US-08-491-954-106		Sequence 106, Appl
9	9	100.0	62	3	US-08-492-076-9		Sequence 9, Appl
10	9	100.0	78	2	US-08-680-326-51		Sequence 51, Appl
11	9	100.0	101	1	US-07-755-859B-14		Sequence 14, Appl
12	9	100.0	101	1	US-08-457-616-14		Sequence 14, Appl
13	9	100.0	299	2	US-08-637-759B-31		Sequence 31, Appl
14	9	100.0	299	2	US-08-871-355A-31		Sequence 31, Appl
15	9	100.0	352	6	5242821-11		Patent No. 5242821
16	9	100.0	454	1	US-08-846-134-2		Sequence 2, Appl
17	9	100.0	492	1	US-08-470-720-3		Sequence 3, Appl
18	9	100.0	573	1	US-08-709-912-18		Sequence 18, Appl
19	9	100.0	573	2	US-09-047-370-18		Sequence 18, Appl
20	9	100.0	599	3	US-08-705-875A-2		Sequence 2, Appl
21	9	100.0	600	5	PCT-US91-02766-24		Sequence 24, Appl
22	9	100.0	628	2	US-08-874-186-45		Sequence 45, Appl
23	9	100.0	660	3	US-09-284-782-15		Sequence 15, Appl
24	9	100.0	663	4	US-08-998-416-187		Sequence 187, Appl
25	9	100.0	706	3	US-08-654-025-4		Sequence 4, Appl
26	9	100.0	713	1	US-08-234-939-9		Sequence 9, Appl
27	9	100.0	713	1	US-08-558-865-9		Sequence 9, Appl

28	9	100.0	717	4	US-08-998-416-849	Sequence 849, Appl
29	9	100.0	719	4	US-08-998-416-1138	Sequence 1138, Appl
30	9	100.0	720	4	US-08-998-416-628	Sequence 628, Appl
31	9	100.0	746	4	US-08-991-789A-261	Sequence 261, Appl
32	9	100.0	836	4	US-09-077-775A-11	Sequence 11, Appl
33	9	100.0	837	1	US-08-832-883-56	Sequence 56, Appl
34	9	100.0	837	2	US-08-832-877-56	Sequence 56, Appl
35	9	100.0	839	3	US-08-817-926-50	Sequence 50, Appl
36	9	100.0	856	4	US-08-998-416-289	Sequence 289, Appl
37	9	100.0	890	1	US-08-234-939-8	Sequence 8, Appl
38	9	100.0	890	1	US-08-558-865-8	Sequence 8, Appl
39	9	100.0	943	3	US-08-705-875A-3	Sequence 3, Appl
40	9	100.0	967	4	US-08-960-780-47	Sequence 47, Appl
41	9	100.0	967	4	US-09-073-898-47	Sequence 47, Appl
42	9	100.0	968	3	US-08-705-875A-9	Sequence 9, Appl
43	9	100.0	972	4	US-08-960-780-48	Sequence 48, Appl
44	9	100.0	972	4	US-09-073-898-48	Sequence 48, Appl
45	9	100.0	975	3	US-09-015-754-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-418-071-15
; Sequence 15, Application US/08418071
; Patent No. 5846705
; GENERAL INFORMATION:
; APPLICANT: Wu, Rey-Yuh
; APPLICANT: You, Li-Ru
; APPLICANT: Soong, Tai-Seng
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: other nucleic acid
; ANTI-SENSE: NO
US-08-418-071-15

Query Match 100.0%; Score 9; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu Oct 25 13:08:59 2001

us-09-462-955-5.rni

; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF

; FILE REFERENCE: 5820.546

; CURRENT APPLICATION NUMBER: US/09/374,493

; CURRENT FILING DATE: 1999-08-13

; EARLIER APPLICATION NUMBER: 09/150,133

; EARLIER FILING DATE: 1998-09-09

; EARLIER APPLICATION NUMBER: 60/072,994

; EARLIER FILING DATE: 1998-01-29

; EARLIER APPLICATION NUMBER: PCT/US99/16750

; EARLIER FILING DATE: 1999-07-23

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)

; SEQ ID NO 33

; LENGTH: 30

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-09-374-493-33

Query Match 100.0%; Score 9; DB 4; Length 30;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9

|||||

Db 4 ctagtattta 12

RESULT 5

US-09-374-824-33

; Sequence 33, Application US/09374824

; Patent No. 6207414

; GENERAL INFORMATION:

; APPLICANT: The Board of Regents of the University of Oklahoma

; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF

; FILE REFERENCE: 5820.547

; CURRENT APPLICATION NUMBER: US/09/374,824

; CURRENT FILING DATE: 1999-08-13

; EARLIER APPLICATION NUMBER: 09/150,133

; EARLIER FILING DATE: 1998-09-09

; EARLIER APPLICATION NUMBER: 60/072,994

; EARLIER FILING DATE: 1998-01-29

; EARLIER APPLICATION NUMBER: PCT/US99/16750

; EARLIER FILING DATE: 1999-07-23

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)

; SEQ ID NO 33

; LENGTH: 30

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-09-374-824-33

Query Match 100.0%; Score 9; DB 4; Length 30;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9

|||||

Db 4 ctagtattta 12

RESULT 6

US-09-374-492-33

; Sequence 33, Application US/09374492

; Patent No. 6207432

; GENERAL INFORMATION:

; APPLICANT: The Board of Regents of the University of Oklahoma

; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF

Qy 1 ctagtattta 9

|||||

Db 2 CTAGTATTTA 10

RESULT 2

US-09-150-133-33

; Sequence 33, Application US/09150133B

; Patent No. 6060295

; GENERAL INFORMATION:

; APPLICANT: The Board of Regents of the University of Oklahoma

; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING

; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF

; FILE REFERENCE: 5820.504

; CURRENT APPLICATION NUMBER: US/09/150,133B

; CURRENT FILING DATE: 1998-09-09

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)

; SEQ ID NO 33

; LENGTH: 30

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-09-150-133-33

Query Match 100.0%; Score 9; DB 3; Length 30;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9

|||||

Db 4 ctagtattta 12

RESULT 3

US-09-150-141-33

; Sequence 33, Application US/09150141B

; Patent No. 6071732

; GENERAL INFORMATION:

; APPLICANT: The Board of Regents of the University of Oklahoma

; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING

; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF

; FILE REFERENCE: 5820.495

; CURRENT APPLICATION NUMBER: US/09/150,141B

; CURRENT FILING DATE: 1998-09-09

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)

; SEQ ID NO 33

; LENGTH: 30

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: primer

US-09-150-141-33

Query Match 100.0%; Score 9; DB 3; Length 30;

Best Local Similarity 100.0%; Pred. No. 1.4e+03;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattta 9

|||||

Db 4 ctagtattta 12

RESULT 4

US-09-374-493-33

; Sequence 33, Application US/09374493

; Patent No. 6204016

; GENERAL INFORMATION:

; APPLICANT: The Board of Regents of the University of Oklahoma

FILE REFERENCE: 5820.545
CURRENT APPLICATION NUMBER: US/09/374,492
CURRENT FILING DATE: 1999-08-13
EARLIER APPLICATION NUMBER: 09/150,141
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 60/072,994
EARLIER FILING DATE: 1998-01-29
EARLIER APPLICATION NUMBER: PCT/US99/16750
EARLIER FILING DATE: 1999-07-23
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
SEQ ID NO 33
LENGTH: 30
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-374-492-33

Query Match 100.0%; Score 9; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9
Db 4 ctagtattta 12

RESULT 7
US-08-817-926-41/c
Sequence 41, Application US/08817926
Patent No. 6001590
GENERAL INFORMATION:
APPLICANT: Kameda, Toshihiro
APPLICANT: Suda, Hisako
APPLICANT: Tamai, Yukio
APPLICANT: Iwamatsu, Akihiro
APPLICANT: Kato, No. 6001590uo
APPLICANT: Sakai, Yasuyoshi
TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINII
TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC-compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,926
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/02597
FILING DATE: 12-SEP-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 234133/1995
FILING DATE: 12-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 42536/1996
FILING DATE: 29-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 081356/0112
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA"
US-08-817-926-41

Query Match 100.0%; Score 9; DB 3; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9
Db 18 CTAGTATTTA 10

RESULT 8
US-08-491-954-106
Sequence 106, Application US/08491954
Patent No. 6096321
GENERAL INFORMATION:
APPLICANT: Girardeau, Jean-Pierre
APPLICANT: Martin, Christine
APPLICANT: Mechin, Marie-Claire
APPLICANT: Der Vartanian, Maurice
APPLICANT: Bousquet, Francois
TITLE OF INVENTION: SUB-UNIT OF CS31A PROTEIN CAPSULE
TITLE OF INVENTION: MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A
TITLE OF INVENTION: PROTEIN CAPSULE INCLUDING SUCH A SUB-UNIT, AND
TITLE OF INVENTION: MICROORGANISMS WHOSE OUTER MEMBRANE CARRIES SUCH
TITLE OF INVENTION: SUB-UNITS, AND PROCEDURE FOR OBTAINING AND UTILIZING SUCH
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: WEISER & ASSOCIATES
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/491,954
FILING DATE: 16-FEB-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/01281
FILING DATE: 21-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6264P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 41 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA

us-09-462-955-5.rni

Thu Oct 25 13:08:59 2001

```

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..41
;
US-08-491-954-106

Query Match      100.0%; Score 9; DB 3; Length 41;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
    |||||
Db 1 CTAGTATTATA 9

RESULT 9
US-08-492-076-9
; Sequence 9, Application US/08492076A
; Patent No. 6060064
; GENERAL INFORMATION:
; APPLICANT: Adams, Sally E.
; APPLICANT: Burus, Nigel R.
; APPLICANT: Richardson, Simon M.
; TITLE OF INVENTION: No. 6060064e1 Proteinaceous Particles
; FILE REFERENCE: 10180.60968
; CURRENT APPLICATION NUMBER: US/08/492.076A
; CURRENT FILING DATE: 1995-06-28
; EARLIER APPLICATION NUMBER: PCT/GB93/02656
; EARLIER FILING DATE: 1993-12-24
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 62
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Partial V3
; OTHER INFORMATION: loop of HIV-1 further comprising restriction
; OTHER INFORMATION: enzyme sites.
US-08-492-076-9

Query Match      100.0%; Score 9; DB 3; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctagtattata 9
    |||||
Db 1 ctagtattata 9

RESULT 10
US-08-680-326-51
; Sequence 51, Application US/08680326
; Patent No. 5925733
; GENERAL INFORMATION:
; APPLICANT: ROSE, TIMOTHY M.
; APPLICANT: BOSCH, MARX
; APPLICANT: STRAND, KURT
; APPLICANT: TODARO, GEORGE J.
; TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
; TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
; TITLE OF INVENTION: FIBROMATOSIS
; NUMBER OF SEQUENCES: 152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
;
US-07-795-859B-14
; Sequence 14, Application US/07795859B
; Patent No. 5422262
; GENERAL INFORMATION:
; APPLICANT: Anderson, Stefan
; APPLICANT: Russell, David W.
; TITLE OF INVENTION: Steroid 5 $\alpha$ -Reductases
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White and Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/795.859B
; FILING DATE: 18-NOV-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:260/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 474-7677
; TELEFAX: (512) 320-7200
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 101 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-07-795-859B-14

```

Query Match 100.0%; Score 9; DB 1; Length 101;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9
 |||||
 Db 26 CTAGTATTA 18

RESULT 12

US-08-457-616-14/c
 ; Sequence 14, Application US/08457616
 ; Patent No. 5679521
 ; GENERAL INFORMATION:
 ; APPLICANT: Anderson, Stefan
 ; APPLICANT: Russell, David W.
 ; TITLE OF INVENTION: Steroid 5'-Reductases
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White and Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII-DOS
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/457,616
 ; FILING DATE: 01-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/795,859
 ; FILING DATE: 18-NOV-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: UTSD:260/PAR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 320-7200
 ; TELEFAX: (512) 474-7677
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 101 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-457-616-14

Query Match 100.0%; Score 9; DB 1; Length 101;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9
 |||||
 Db 26 CTAGTATTA 18

RESULT 13

US-08-637-759B-31/c
 ; Sequence 31, Application US/08637759B
 ; Patent No. 5876931
 ; GENERAL INFORMATION:
 ; APPLICANT: David William Holden
 ; TITLE OF INVENTION: Identification of Genes
 ; NUMBER OF SEQUENCES: 501

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Patrea L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; STREET: 1201 West Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/637,759B
 ; FILING DATE: 03-MAY-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/GB95/02875
 ; FILING DATE: 11-DEC-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pabst, Patrea L.
 ; REGISTRATION NUMBER: 31,284
 ; REFERENCE/DOCKET NUMBER: RPLS 101
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (404) 873-8794
 ; TELEFAX: (404) 873-8795
 ; INFORMATION FOR SEQ ID NO: 31:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 299 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Partial sequence of Salmonella typhimurium
 ; ORGANISM: virulence gene
 ; US-08-637-759B-31

Query Match 100.0%; Score 9; DB 2; Length 299;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9
 |||||
 Db 52 CTAGTATTA 44

RESULT 14

US-08-871-355A-31/c
 ; Sequence 31, Application US/08871355A
 ; Patent No. 6015669
 ; GENERAL INFORMATION:
 ; APPLICANT: David William Holden
 ; TITLE OF INVENTION: Identification of Genes
 ; NUMBER OF SEQUENCES: 501
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Patrea L. Pabst
 ; STREET: 2800 One Atlantic Center
 ; STREET: 1201 West Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: USA
 ; ZIP: 30309-3450
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

Thu Oct 25 13:08:59 2001

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
; US-08-871-355A-31

```

```

Query Match      100.0%; Score 9; DB 3; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ctagtattata 9
    |||
DB 52 CTAGTATTA 44

```

```

RESULT 15
5242821-11
; Patent No. 5242821
; APPLICANT: PALVA, LIKKA;SIBAKHOV, MERVI
; TITLE OF INVENTION: LACTOCOCCUS PROMOTER AND SIGNAL
; SEQUENCES FOR EXPRESSION IN BACTERIA
; NUMBER OF SEQUENCES: 27
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/377,450
; FILING DATE: 10-JUL-1989
; SEQ ID NO:11:
; LENGTH: 352
5242821-11

```

```

Query Match      100.0%; Score 9; DB 6; Length 352;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ctagtattata 9
    |||
DB 131 ctagtattata 139

```

```

Search completed: October 24, 2001, 10:00:31
Job time: 216 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:59 ; Search time 5479.82 Seconds
(without alignments)
15.525 Million cell updates/sec

Title: US-09-462-955-5
Perfect score: 9
Sequence: 1 ctagtattata 9

Scoring table: IDENTITY_NUC-
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estrol1:*
83: em_estrol2:*
84: em_estrol3:*
85: em_estrol4:*
86: em_estrol5:*
87: em_estrol6:*
88: em_estrol7:*
89: em_estrol8:*
90: em_estrol9:*
91: em_estrol10:*
92: em_estrol11:*
93: em_estrol12:*
94: em_estrol13:*
95: em_estrol14:*
96: em_estrol15:*
97: em_estrol16:*
98: em_estrol17:*
99: em_estrol18:*
100: em_estrol19:*
101: em_estrol20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

Thu Oct 25 13:09:00 2001

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
```

```
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	100.0	40	167	BE386120 601277233
2	9	100.0	55	249	AZ769245 1M0569010
3	9	100.0	57	143	AF097417 AF097417
4	9	100.0	57	243	AZ434302 1M0220E09
5	9	100.0	58	10	AA637860 VR29F01.r
6	9	100.0	59	113	AW215812 u099e06.x
7	9	100.0	59	113	AW215820 u099f06.x
8	9	100.0	62	258	TA82F100 T. brucei
9	9	100.0	66	250	AZ808515 2M0072A07
10	9	100.0	67	24	A1761414 w65c06.x
11	9	100.0	71	4	AA250291 m261g02.r
12	9	100.0	71	123	AW989354 u25b07.y
13	9	100.0	71	138	BE627067 u15f06.y
14	9	100.0	71	138	BE630111 u15f06.x
15	9	100.0	71	138	BE688187 u43b06.x
16	9	100.0	71	138	BE692587 u43b06.y
17	9	100.0	71	141	BE849664 u087f04.y
18	9	100.0	73	167	BE448984 ut50h06.y
19	9	100.0	75	2	AA079298 zm97g12.s
20	9	100.0	78	113	AW226900 u063c01.y
21	9	100.0	80	138	BE627009 u08h09.y
22	9	100.0	84	2	AA079297 zm97g12.r
23	9	100.0	84	106	AU012912 AU012912
24	9	100.0	85	13	AA937702 o104g12.s
25	9	100.0	87	11	AA760419 v76h02.r
26	9	100.0	87	249	AZ783472 2M0025B18
27	9	100.0	90	155	C01576 HUMG000858
28	9	100.0	93	159	N28044 EST000258 S
29	9	100.0	94	20	A1465178 mv98a02.x
30	9	100.0	94	241	AZ309606 1M0016B06
31	9	100.0	94	247	AZ660200 1M0538I08
32	9	100.0	95	31	AV562434 AV562434
33	9	100.0	97	7	AA419984 vF38h02.r
34	9	100.0	99	107	AU014567 AU014567
35	9	100.0	100	10	AA689650 vS10g11.r
36	9	100.0	100	24	A1718783 as58h08.x
37	9	100.0	101	164	BE152001 Q13-HT031
38	9	100.0	101	164	BE179567 I13-HT061
39	9	100.0	102	186	BE365112 P13_24_B0
40	9	100.0	102	189	T61718 yb92e07.r1
41	9	100.0	102	237	AZ059959 RPCI-23-4
42	9	100.0	104	8	AA518125 v123g11.r
43	9	100.0	104	121	AW810470 MR4-ST012
44	9	100.0	104	164	BE145670 I10-HT020
45	9	100.0	105	1	AA014683 mh09g05.r

ALIGNMENTS

Result No.	Score	Query Match	Length	DB ID	Description
1	9	100.0	40	bp	EST
2	9	100.0	40	bp	EST
3	9	100.0	40	bp	EST
4	9	100.0	40	bp	EST
5	9	100.0	40	bp	EST
6	9	100.0	40	bp	EST
7	9	100.0	40	bp	EST
8	9	100.0	40	bp	EST
9	9	100.0	40	bp	EST
10	9	100.0	40	bp	EST
11	9	100.0	40	bp	EST
12	9	100.0	40	bp	EST
13	9	100.0	40	bp	EST
14	9	100.0	40	bp	EST
15	9	100.0	40	bp	EST
16	9	100.0	40	bp	EST
17	9	100.0	40	bp	EST
18	9	100.0	40	bp	EST
19	9	100.0	40	bp	EST
20	9	100.0	40	bp	EST
21	9	100.0	40	bp	EST
22	9	100.0	40	bp	EST
23	9	100.0	40	bp	EST
24	9	100.0	40	bp	EST
25	9	100.0	40	bp	EST
26	9	100.0	40	bp	EST
27	9	100.0	40	bp	EST
28	9	100.0	40	bp	EST
29	9	100.0	40	bp	EST
30	9	100.0	40	bp	EST
31	9	100.0	40	bp	EST
32	9	100.0	40	bp	EST
33	9	100.0	40	bp	EST
34	9	100.0	40	bp	EST
35	9	100.0	40	bp	EST
36	9	100.0	40	bp	EST
37	9	100.0	40	bp	EST
38	9	100.0	40	bp	EST
39	9	100.0	40	bp	EST
40	9	100.0	40	bp	EST
41	9	100.0	40	bp	EST
42	9	100.0	40	bp	EST
43	9	100.0	40	bp	EST
44	9	100.0	40	bp	EST
45	9	100.0	40	bp	EST

Result No. 1
 BE386120
 LOCUS
 DEFINITION
 601277233F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3617961 5', mRNA sequence.
 BE386120
 BE386120
 EST.
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 40)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/BTP
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LICM286 row: b column: 10.
 Location/Qualifiers
 source
 1..40
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3617961"
 /clone_lib="NIH_MGC_20"
 /tissue_type="melanotic melanoma"
 /lab_host="DPH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(S). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES

BASE COUNT 12 a 8 c 11 g 9 t
 ORIGIN
 Query Match 100.0%; Score 9; DB 167; Length 40;
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES

QY 1 ctagtattata 9
 Db 11 CTAGTATTATA 19
 RESULT 2
 LOCUS
 DEFINITION
 A2769245 55 bp DNA GSS 16-FEB-2001
 1M0569010R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M 569010 R, DNA sequence.
 A2769245
 A2769245.1 GI:12889180
 GSS.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 55)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0569 row: C column: 10
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 55.
 Location/Qualifiers
 1..55
 /organism="Mus musculus"

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9
|||||
Db 50 CTAGTATTTA 42

RESULT 4

AZ434302

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BASE COUNT

ORIGIN

33 a 4 c 4 g 14 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0220E09"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans"

/db_xref="taxon:105273"

/clone_lib="Vespertilio superans library"

/note="country=South Korea; Kangwon Province"

BASE COUNT

ORIGIN

19 a 9 c 9 g 20 t

Query Match

Best Local Similarity

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattta 9

Db 18 CTAGTATTTA 10

RESULT 3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 57

/organism="Vespertilio superans

2y 1 ct

Thu Oct 25 13:09:00 2001

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps@remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

FEATURES source
 QY 1 ctagtattata 9
 Db 5 CTAGTATTA 13

BASE COUNT 14 a 11 c 7 g 27 t
ORIGIN Mus musculus
 Query Match 100.0%; Score 9; DB 113; Length 59;
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ctagtattata 9
 Db 5 CTAGTATTA 13

RESULT 8
TA82F10Q/c 62 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 82F10, reverse sequence,
DEFINITION genomic survey sequence.
ACCESSION AL459971
VERSION AL459971.1 GI:11860296
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.

REFERENCE 1 (bases 1 to 62)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
 Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
 Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
 nhlsanger.ac.uk

COMMENT Constructed at the Institute for Genomic Research (TIGR),
 Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
 to give a tight size distribution (4 kb). The v + i method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES source
 Location/Qualifiers
 1..62
 /organism="Trypanosoma brucei"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="82f10"
 27 a 8 c 11 g 16 t
BASE COUNT 27 a 8 c 11 g 16 t
ORIGIN

Query Match 100.0%; Score 9; DB 258; Length 62;
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtattata 9
 Db 18 CTAGTATTA 10

RESULT 9
AZ808515 66 bp DNA GSS 20-FEB-2001
LOCUS 2M0072A07F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0072A07 F, DNA sequence.
ACCESSION AZ808515
VERSION AZ808515.1 GI:12973933
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 66)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
 and Wright, D., Weis, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0072 row: A column: 07
 Seq primer: CGTGTAAACGACGCCAGT
 Class: plasmid ends
 High quality sequence stop: 66.
 Location/Qualifiers
 1..66
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0072A07"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The

Thu Oct 25 13:09:00 2001

```

QY 1 ctagtatta 9
Db 66 CTAGTATTA 58

RESULT 12
AW989354/c 71 bp mRNA EST 02-JUN-2000
LOCUS u15f06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1512373 5', mRNA sequence.
ACCESSION AW989354
VERSION AW989354.1 GI:8184733
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1081639
Seq primer: -40RP from Gibco.
Location/Qualifiers
1..71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3372035"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 32 a 10 c 12 g 17 t
ORIGIN
1..71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1512373"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
Query Match 100.0%; Score 9; DB 123; Length 71;
Best Local Similarity 100.0%; Pred. NO. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtatta 9
Db 66 CTAGTATTA 58

RESULT 14
BE630111 71 bp mRNA EST 25-AUG-2000
LOCUS u15f06.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:3372035 3', mRNA sequence.
ACCESSION BE630111
VERSION BE630111.1 GI:9912799
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: u15f06.y2
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1081639.
Location/Qualifiers
1..71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3372035"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 32 a 10 c 12 g 17 t
ORIGIN
1..71
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1512373"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
Query Match 100.0%; Score 9; DB 123; Length 71;
Best Local Similarity 100.0%; Pred. NO. 3.2e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctagtatta 9
Db 66 CTAGTATTA 58

RESULT 13
BE627067/c 71 bp mRNA EST 24-AUG-2000
LOCUS u15f06.y2 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:3372035 5', mRNA sequence.
ACCESSION BE627067
VERSION BE627067.1 GI:9907487
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS

```

gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 17 a 12 c 11 g 31 t
 ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctagtattta 9
 |||||
 Db 5 CTAGTATTA 13

RESULT 15
 BE688187
 LOCUS
 DEFINITION 71 bp mRNA EST 11-SEP-2000
 uw43b06.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
 IMAGE:3464723 3', mRNA sequence.
 BE688187
 BE688187
 BE688187.1 GI:10075811
 EST.
 house mouse.
 Mus musculus.
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 71)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:1384083.

FEATURES
 source
 1..71
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:3464723"
 /clone_lib="Soares_mammary_gland_NMLMG"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library is normalized. Library
 was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 18 a 12 c 9 g 32 t
 ORIGIN

Query Match 100.0%; Score 9; DB 138; Length 71;
 Best Local Similarity 100.0%; Pred. No. 3.2e+04;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctagtattta 9
 |||||
 Db 5 CTAGTATTA 13

us-09-462-955-5.rst

Thu Oct 25 13:09:00 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:07:28 ; Search time 393.52 Seconds
(without alignments)
36.699 Million cell updates/sec

Title: US-09-462-955-6
Perfect score: 23
Sequence: 1 ctgccagggccagagcctaggga 23

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDSL1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	44	AA02689	CFDV-DNA fragment
2	23	100.0	44	AA02687	CFDV-DNA fragment
3	19.8	86.1	324	AA025325	Human ORF80
4	19.8	86.1	1592	AA027241	Human secreted pro
5	18.8	81.7	524	AA049239	Human secreted pro
6	18.2	79.1	2512	AA012867	Human secreted pro
7	18.2	79.1	3131	AA012867	JM1-229 cell line
8	17.8	77.4	496	AA01708	Gene 036 different
9	17.8	77.4	1658	AA038123	Human FATP alterna
10	17.8	77.4	2219	AA018209	Lung cancer associ
11	17.8	77.4	2222	AA038122	Human FATP protein
				AA038124	Human FATP variant

C 12	17.8	77.4	2816	19	AA006641	cDNA encoding huma
C 13	17.8	77.4	2818	21	AA0294053	DNA encoding human
C 14	17.8	77.4	3217	20	AA000365	Nucleotide sequenc
C 15	17.8	77.4	3694	20	AA000356	Nucleotide sequenc
C 16	17.8	77.4	3704	20	AA000352	Nucleotide sequenc
C 17	17.8	77.4	6744	20	AA038125	Human FATP genomic
C 18	17.2	74.8	1851	20	AA027845	Mutant lipase codi
C 19	17.2	74.8	5349	21	AA065527	Porcine BAC-PfGF2-
C 20	16.8	73.0	534720	19	AA030458	Rhizobium species
C 21	16.8	73.0	536165	19	AA030459	Rhizobium species
C 22	16.6	72.2	410	21	AA002239	Human secreted pro
C 23	16.6	72.2	446	21	AA009796	Human secreted pro
C 24	16.6	72.2	1662	20	AA038080	Human forkhead tra
C 25	16.6	72.2	1662	20	AA031672	Human forkhead tra
C 26	16.6	72.2	1845	20	AA033613	Human FKHL7 coding
C 27	16.6	72.2	1856	21	AA076987	Human breast tumou
C 28	16.6	72.2	2828	21	AA014001	DNA encoding human
C 29	16.6	72.2	3148	21	AA076239	Human ORF ORF2542
C 30	16.6	72.2	3946	20	AA038079	Human forkhead tra
C 31	16.6	72.2	3946	20	AA031671	Human forkhead tra
C 32	16.4	71.3	300	20	AA015019	Human FKHL7 coding
C 33	16.4	71.3	846	20	AA011180	Human gene express
C 34	16.4	71.3	1210	20	AA052236	DNA encoding trans
C 35	16.4	71.3	1210	21	AA077549	Protein PRO224 cDN
C 36	16.4	71.3	1210	21	AA049721	Human PRO224 cDNA
C 37	16.4	71.3	1210	22	AA072394	Human PRO224 cDNA
C 38	16.4	71.3	1210	22	AA097425	Human PRO224 cDNA
C 39	16.4	71.3	1264	21	AA053830	Human angiogenesis
C 40	16.4	71.3	1270	20	AA011187	Sequence encoding
C 41	16.4	71.3	2206	15	AA074445	Gene encoding tran
C 42	16.4	71.3	2945	16	AA087790	DNA located 3' of
C 43	16.4	71.3	3690	21	AA021281	S. lavenulae mcr
C 44	16.4	71.3	3690	21	AA035159	Human low adenosin
C 45	16.4	71.3	4077	19	AA028999	Human adenosine re
						Human endothelial

ALIGNMENTS

RESULT 1
ID AA02689 standard; DNA; 44 BP.
XX
AC AA02689;
XX
DT 10-MAY-1999 (first entry)
XX
DE CFDV-DNA fragment stem loop repeat region.
XX
KW Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;
KW Yeast promoter; tissue-specific; phloem-specific; CamV 35S promoter; ss.
XX
OS Coconut foliar decay virus
XX
FH Key Location/Qualifiers
FT stem_loop 4..22
FT /*tag= a
FT /*note= "RPT1"
FT misc_binding 4..10
FT /*tag= b
FT /*note= "Region binds to nucleotides 16 to 22"
FT misc_binding 16..22
FT /*tag= c
FT /*note= "Region binds to nucleotides 4 to 10"
FT stem_loop 27..43
FT /*tag= d
FT /*note= "RPT2"
FT misc_binding 27..32
FT /*tag= e
FT /*note= "Region binds to nucleotides 38 to 43"
FT misc_binding 36..43
FT /*tag= f
FT /*note= "Region binds to nucleotides 27 to 32"

FT DE19730502-A1. /*tag= f /note= "Region binds to nucleotides 27 to 32"

XX 21-JAN-1999.

XX 16-JUL-1997; 97DE-1030502.

XX 16-JUL-1997; 97DE-1030502.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;

XX WPI; 1999-096863/09.

XX Coconut foliar decay virus promoters - for gene expression in

XX bacteria and yeasts

XX Disclosure; Fig 4; 14pp; German.

XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment

XX that includes the stem-loop structure of CFDV DNA but lacks the

XX translation start codons of open reading frames ORF1 and/or ORF2. The

XX new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a

XX promoter for tissue-specific (especially phloem-specific) gene expression

XX in plants and for production of chimeric constructs for transient or

XX stable expression. Certain fragments of CFDV DNA have stronger promoter

XX activity in E. coli than the CaMV 35S promoter.

XX Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;

XX Query Match 100.0%; Score 23; DB 20; Length 44;

XX Best Local Similarity 100.0%; Pred. No. 3.5;

XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgccaggccgaagcctggga 23

DB 1 ctgccaggccgaagcctggga 23

RESULT 2

AA02687

ID AAX02687 standard; DNA; 44 BP.

XX AC AAX02687;

XX 10-MAY-1999 (first entry)

XX CFDV-DNA fragment stem loop repeat region.

XX Stem loop; coconut foliar decay virus; CFDV; bacterial promoter;

XX yeast promoter; tissue-specific; phloem-specific; CaMV 35S promoter; ss.

XX Coconut foliar decay virus

XX Key Location/Qualifiers

FT stem_loop 4..22 /*tag= a

FT /*note= "RPT1"

FT misc_binding 4..10 /*tag= b

FT /*note= "Region binds to nucleotides 16 to 22"

FT misc_binding 16..22 /*tag= c

FT /*note= "Region binds to nucleotides 4 to 10"

FT stem_loop 27..43 /*tag= d

FT /*note= "RPT2"

FT misc_binding 27..32 /*tag= e

FT /*note= "Region binds to nucleotides 38 to 43"

FT misc_binding 38..43

FT

FT DE19730535-A1.

XX 21-JAN-1999.

XX 16-JUL-1997; 97DE-1030535.

XX 16-JUL-1997; 97DE-1030535.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Becker D, Hehn A, Randles JW, Rohde W, Salamini F;

XX WPI; 1999-096867/09.

XX Coconut foliar decay virus promoters - for gene expression in

XX plants, bacteria and yeasts

XX Disclosure; Fig 4; 12pp; German.

XX This invention describes a coconut foliar decay virus (CFDV) DNA fragment

XX that includes the stem-loop structure of CFDV DNA but lacks the

XX translation start codons of open reading frames ORF1 and/or ORF2. The

XX new CFDV DNA fragment is useful as a bacterial or yeast promoter, as a

XX promoter for tissue-specific (especially phloem-specific) gene expression

XX in plants and for production of chimeric constructs for transient or

XX stable expression. Certain fragments of CFDV DNA have stronger promoter

XX activity in E. coli than the CaMV 35S promoter.

XX Sequence 44 BP; 8 A; 16 C; 17 G; 3 T; 0 other;

XX Query Match 100.0%; Score 23; DB 20; Length 44;

XX Best Local Similarity 100.0%; Pred. No. 3.5;

XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgccaggccgaagcctggga 23

DB 1 ctgccaggccgaagcctggga 23

RESULT 3

AA075325/c

ID AAC75325 standard; CDNA; 324 BP.

XX AC AAC75325;

XX 08-FEB-2001 (first entry)

XX Human ORF80 polynucleotide sequence SEQ ID NO:1759.

XX Human; open reading frame; ORF8; detection; cytostatic; hepatotropic;

XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

XX hypotensive; dermatological; immunosuppressive; antiinflammatory;

XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

XX neurodegenerative disorder; osteoarthritis; graft vs host disease;

XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

XX cholesterol ester storage; systemic lupus erythematosus; infection;

XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

XX bone damage; cartilage damage; antiinflammatory disease; coagulation;

XX thrombosis; contraceptive; ss.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX

```

XX PF 31-MAR-2000; 2000WO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX XX Shimkets RA, Leach M;
XX PI WPI: 2000-602362/57.
XX XX P-PSDB; AAB41116.
XX DR Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS Claim 5; Page 1386-1387; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX CC antithyroid; and antianemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy.
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, cholesterol ester storage, systemic lupus
XX CC erythematosus, malaria, autoimmune disorders, asthma,
XX CC allergic or aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX SQ Sequence 324 BP; 82 A; 84 C; 89 G; 69 T; 0 other;

Query Match 86.1%; Score 19.8; DB 21; Length 324;
Best Local Similarity 91.3%; Pred. No. 46;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaaggcctggga 23
DB 71 CTGCCAGGCCGAGTGCTGGGA 49

RESULT 4
AAZ27241/c
ID AAZ27241 standard; cDNA; 1592 BP.
XX AAZ27241;
XX AC AAZ27241;
XX DT 23-NOV-1999 (first entry)
XX XX Human secreted protein cDNA encoding gene 9.
XX DE Secreted protein; human; gene therapy; diagnosis; treatment; cancer;
XX KW protein therapy; tumor; neurodegenerative disorder; blood disorder; AIDS;
XX KW developmental abnormality; leukemia; immune system; autoimmune disease;
XX KW hepatic disease; renal disease; inflammation; allergy; schizophrenia;
XX KW Alzheimer's disease; cognitive disorder; arthritis; infection; psoriasis;
XX KW transplant rejection; diabetes; asthma; sepsis; acne; metabolic disorder;
XX KW cardiovascular disorder; food additive; preservative; ss.

Query Match 86.1%; Score 19.8; DB 20; Length 1592;
Best Local Similarity 91.3%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaaggcctggga 23
DB 1066 CTGCCAGGCCGAGGCTGAGA 1044

RESULT 5
AAZ49239/c
ID AAZ49239 standard; cDNA; 524 BP.
XX AAZ49239;
XX AC AAZ49239;
XX DT 07-MAR-2000 (first entry)
XX XX Human hydrolase homologue HHH-6 cDNA fragment 1514414P6.
XX DE Hydrolase; homologue; HHH-6; N-terminal asparagine amidohydrolase;
XX KW vanin-1; glycosyl hydrolase; glucosylase;
XX KW N-acetylglucosamine 6-P deacetylase; diagnosis;
XX KW treatment; prevention; expression; disorder;
XX KW antagonist; reproductive disorder; cell proliferation; ds.

```

```

OS Homo sapiens.
XX PN WO9946289-A1.
XX XX 16-SEP-1999.
XX PD 11-MAR-1999; 99WO-US05721.
XX PF 12-MAR-1998; 98US-0077686.
XX PR 12-MAR-1998; 98US-0077687.
XX PR 12-MAR-1998; 98US-0077696.
XX PR 12-MAR-1998; 98US-0077714.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX Ruben SM, Ferrie AM, Rosen CA, Florence C, Young PE, Yu G, Ni J;
XX PI WPI: 1999-551363/46.
XX XX P-PSDB; AAY45265.
XX DR New isolated human genes, useful for diagnosis and treatment of, e.g.
XX DR cancers -
XX PT Claim 1a; 155; 306pp; English.
XX PS This invention describes novel human genes and the secreted proteins
XX CC they encode. The polynucleotides and their corresponding secreted
XX CC polypeptides are useful for preventing, treating or ameliorating medical
XX CC conditions e.g. by protein or gene therapy. Also pathological conditions
XX CC can be diagnosed by determining the amount of the new polypeptides in a
XX CC sample or by determining the presence of mutations in the new
XX CC polynucleotides. Specific uses are described for each of the
XX CC polynucleotides of the invention, based on which tissues they are most
XX CC highly expressed in, and include developing products for the diagnosis or
XX CC treatment of cancer, tumors, neurodegenerative disorders, developmental
XX CC abnormalities, blood disorders, leukemias, diseases of the immune system,
XX CC autoimmune diseases, hepatic and renal disease, inflammation, allergies,
XX CC Alzheimer's and cognitive disorders, schizophrenia, arthritis,
XX CC infections, AIDS, transplant rejection, diabetes, asthma, sepsis, acne,
XX CC psoriasis, cardiovascular disorders, and metabolic disorders. The
XX CC polypeptides or polynucleotides can also be used as food additives or
XX CC preservatives. The polypeptides are also useful for identifying their
XX CC binding partners. AAZ27233-227265 encode human secreted proteins
XX CC described in the method of the invention.
XX SQ Sequence 1592 BP; 432 A; 382 C; 418 G; 360 T; 0 other;

Query Match 86.1%; Score 19.8; DB 20; Length 1592;
Best Local Similarity 91.3%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccaggccgaaggcctggga 23
DB 1066 CTGCCAGGCCGAGGCTGAGA 1044

RESULT 5
AAZ49239/c
ID AAZ49239 standard; cDNA; 524 BP.
XX AAZ49239;
XX AC AAZ49239;
XX DT 07-MAR-2000 (first entry)
XX XX Human hydrolase homologue HHH-6 cDNA fragment 1514414P6.
XX DE Hydrolase; homologue; HHH-6; N-terminal asparagine amidohydrolase;
XX KW vanin-1; glycosyl hydrolase; glucosylase;
XX KW N-acetylglucosamine 6-P deacetylase; diagnosis;
XX KW treatment; prevention; expression; disorder;
XX KW antagonist; reproductive disorder; cell proliferation; ds.

```

OS Synthetic.
 XX Homo sapiens.
 XX W09961626-A2.
 XX
 XX 02-DEC-1999.
 XX
 XX 28-MAY-1999; 99WO-US12021.
 XX
 XX 29-MAY-1998; 98US-0087236.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Bandman O, Hillman JL, Yue H, Lal P, Corley NC, Guegler KJ;
 PI Patterson C, Baughn MR;
 XX
 XX WPI; 2000-062716/05.
 XX
 XX New human hydrolase homologues, useful for treating or preventing a
 PT carbohydrate metabolism disorder -
 PT
 PT
 PT
 PS Disclosure; Page 77; 91pp; English.
 XX
 XX Sequences AA249235, AA249237, AA249239, AA249242 and AA249259 represent
 CC human hydrolase homologue HHH-6 cDNA fragments. The full-length
 CC cDNA sequence is given in AA249228. Nucleic acids encoding HHH-6
 CC were initially identified in a uterine endometrium tissue cDNA
 CC library. HHH-6 has homology to N-acetylglucosamine 6-P
 CC deacetylase. The invention relates to human hydrolase
 CC homologues HHH-1 to HHH-7 (AA249235-249259) which respectively
 CC have homology to N-terminal asparagine amidohydrolase, vanin-1,
 CC glycosyl hydrolases, glucohydrolase and N-acetylglucosamine 6-P
 CC deacetylase. Such homologues are useful in methods for diagnosing,
 CC treating or preventing disorders associated with expression of
 CC hydrolases. The hydrolase homologues are useful for treating or
 CC preventing a carbohydrate metabolism disorder. Antagonists of these
 CC hydrolases can be used to treat or prevent a reproductive or cell
 CC proliferation disorder.
 XX
 XX Sequence 524 BP; 95 A; 160 C; 167 G; 90 T; 12 other;
 SQ
 Query Match 81.7%; Score 18.8; DB 21; Length 524;
 Best Local Similarity 90.9%; Pred. No. 1e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 tgcccaggccgaaggcctggga 23
 Db 299 TGCCAGGCGCCAAAGCCAGGGA 278
 RESULT 6
 AAQ12867/c
 ID AAQ12867 standard; cDNA; 2512 BP.
 XX
 XX AAQ12867;
 XX
 XX 15-OCT-1991 (first entry)
 XX
 XX JMI-229 cell line t-PA.
 XX
 XX Thrombolytic agent; pulmonary embolism; ss.
 XX
 XX Rattus rattus.
 XX
 XX Key Location/Qualifiers
 FT CDS 91..1770
 FT /tag= a
 FT /product= t-PA
 FT sig_peptide 91..186
 FT /tag= b
 FT mat_peptide 187..1770
 FT /tag= c

misc_feature 1228..1230
 /tag= d
 /note= "Lys -> Glu change"
 1921
 misc_feature
 /tag= e
 /note= "A -> C change"
 2475..2480
 /tag= f
 polyA_signal
 2505..2512
 /tag= g
 polyA_site
 W09110447-A.
 XX
 XX 25-JUL-1991.
 PD
 XX 09-JAN-1991; 91WO-GB00025.
 PF
 XX 11-JAN-1990; 90GB-0000629.
 PR
 XX (PORT-) PORTON PROD LTD.
 PA (PUBL-) PUBLIC HEALTH LAB SERVICE.
 PA (UNLO-) UNIV COLLEGE LONDON.
 XX
 XX Atkinson A, Doyle A, Griffiths J, Electriwala A, Kearns M;
 PI Mellings J, North JR, Riley PA, Scawen MD, Small IS, Sutton PM;
 XX
 XX WPI; 1991-237801/32.
 DR P-PSDB; AAR13263.
 XX
 XX Tissue plasminogen activator derived from cell line JMI-229 - for
 PT treatment of blood vessel obstructions by fibrin clots e.g.
 PT myocardial infarction and stroke etc.
 XX
 XX Claim 1; Page 27; 55pp; English.
 XX
 XX The nucleotide sequence shown was elucidated from cDNA and is that
 CC of clone PPA15. The t-PA from the JMI-229 cell line differs from
 CC that of normal rat t-PA, having Serine as the N-terminal amino acid
 CC and Glutamic acid at position 348 counted from the N-terminal of the
 CC molecule. The other nucleotide difference is in the non-coding
 CC region of the cDNA sequence, where cytosine replaces adenine at
 CC position 1921. This may result in alterations to mRNA stability
 CC or translatability.
 CC
 XX Sequence 2512 BP; 680 A; 632 C; 638 G; 562 T; 0 other;
 SQ
 Query Match 79.1%; Score 18.2; DB 12; Length 2512;
 Best Local Similarity 87.0%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ctgcccaggccgaaggcctggga 23
 Db 872 CTGCCAGGCGCCAAAGTCCTGGGA 850
 RESULT 7
 AAAT91708
 ID AAAT91708 standard; cDNA; 3131 BP.
 XX
 XX AAAT91708;
 XX
 XX 11-MAR-1998 (first entry)
 DT
 XX
 XX DE Gene 036 differentially expressed in colon tumour cells.
 XX
 XX Colon tumour; colon cancer; differential expression; gene 036;
 KW human; diagnosis; gene therapy; tumour suppressor;
 KW fingerprint gene; ss.
 XX
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH

(ROSE/) ROSEN C. A.
Ruben SM;
WPI: 2000-587514/55.
P-PSDB: AAB58333.
Lung cancer associated gene sequences, referred to as lung cancer antigens, useful for treatment, prevention, and diagnosis of disorders such as lung cancer -
Claim 1; Page 689-690; 1425pp; English.
polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer associated proteins represented in AAB58106 - AAB58548. Lung cancer associated proteins and polynucleotide sequences, their agonists, and antagonists may have neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular; active general; vulnarary; gastrointestinal general; nephrotropic; antifinfective; gynecological; or antibacterial activity. The invention also includes antibodies specific for the protein or polynucleotide sequences. The lung cancer associated polynucleotide sequences may be used for detection of lung cancer, chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The proteins may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders. The proteins may also be used in the treatment of wounds and infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are used in the course of the invention for the identification and characterisation of the polynucleotide and protein sequences.
Sequence 1658 BP; 386 A; 429 C; 374 G; 463 T; 6 other;
Query Match 77.4%; Score 17.8; DB 21; Length 1658;
Best Local Similarity 90.5%; Pred. NO. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccaggccgaagcctgg 21
DB 674 CTGCCAGGTAGAGGCCTGG 654
RESULT 10
AAZ38122/c
ID AAZ38122 standard; cDNA; 2219 BP.
XX AC AAZ38122;
XX DT 08-FEB-2000 (first entry)
XX DE Human FATP protein encoding cDNA.
XX DE Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;
XX KW long-chain fatty acid metabolism; obesity; human; ss.
XX OS Homo sapiens.
XX PN WO9951740-A2.
XX PD 14-OCT-1999.
XX PF 02-APR-1999; 99WO-EF02295.
XX PR 06-APR-1998; 98EP-0400823.
XX PA (JANC) JANSSEN PHARM NV.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Martin G, Nemoto M, Deeb SS, Auwerx J;
XX WPI: 1999-620202/53.
XX New human fatty acid transport protein, hFATP, useful to screen for inhibitors or enhancers useful to regulate fatty acid metabolism -
Claim 2; Fig 3; 83pp; English.
The invention provides a human fatty acid transport protein (hFATP). hFATP is believed to be involved in the modulation long-chain fatty acid metabolism; the protein and polynucleotides therefore enable production of compositions comprising a component regulating (inhibiting or enhancing) expression of the hFATP gene useful therapeutically to alter intracellular or blood levels of long chain fatty acids. Such compounds are especially useful to treat conditions associated with deficient regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or diabetes or an enhancer to treat obesity. The polynucleotides are also useful to screen compounds for their effects on hFATP expression, e.g. by measuring mRNA transcription in cells/cell extracts (e.g. liver cells) contacted with the compound and comparing with that in non-contacted cells. Sequences AAZ38122-123 represent nucleotide sequence of clones containing 2 kb sequences obtained from a cDNA library of human adipose tissue. The present sequence represents a cDNA encoding the hFATP protein.
Sequence 2219 BP; 369 A; 710 C; 731 G; 406 T; 3 other;
Query Match 77.4%; Score 17.8; DB 20; Length 2219;
Best Local Similarity 90.5%; Pred. NO. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 tgcccaggccgaagcctgg 22
DB 539 TGCCCAGGCAGAGGCAGGG 519
RESULT 11
AAZ38124/c
ID AAZ38124 standard; cDNA; 2222 BP.
XX AC AAZ38124;
XX DT 08-FEB-2000 (first entry)
XX DE Human FATP variant nucleotide sequence.
XX DE Fatty acid transport protein; FATP; hFATP; cardiomyopathy; diabetes;
XX KW long-chain fatty acid metabolism; obesity; human; variant; ss.
XX OS Homo sapiens.
XX PN WO9951740-A2.
XX PD 14-OCT-1999.
XX PF 02-APR-1999; 99WO-EF02295.
XX PR 06-APR-1998; 98EP-0400823.
XX PA (JANC) JANSSEN PHARM NV.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Martin G, Nemoto M, Deeb SS, Auwerx J;
XX WPI: 1999-620202/53.
XX New human fatty acid transport protein, hFATP, useful to screen for inhibitors or enhancers useful to regulate fatty acid metabolism -
Claim 2; Fig 3; 83pp; English.
The invention provides a human fatty acid transport protein (hFATP). hFATP is believed to be involved in the modulation long-chain fatty acid

CC metabolism; the protein and polynucleotides therefore enable production
 CC of compositions comprising a component regulating (inhibiting or
 CC enhancing) expression of the hFATP gene useful therapeutically to alter
 CC intracellular or blood levels of long chain fatty acids. Such compounds
 CC are especially useful to treat conditions associated with deficient
 CC regulation (e.g. may comprise an inhibitor to treat cardiomyopathies or
 CC diabetes or an enhancer to treat obesity. The polynucleotides are also
 CC useful to screen compounds for their effects on hFATP expression, e.g.
 CC by measuring mRNA transcription in cells/cell extracts (e.g. liver
 CC cells) contacted with the compound and comparing with that in non-
 CC contacted cells. The present sequence represents a variant of the
 CC nucleotide sequence encoding the hFATP protein, obtained from a cDNA
 CC library of human adipose tissue.
 XX
 SQ Sequence 2222 BP; 393 A; 666 C; 696 G; 464 T; 3 other;

Query Match 77.4%; Score 17.8; DB 20; Length 2222;
 Best Local Similarity 90.5%; Pred. No. 2.2e-02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgccagggccgaagcctggg 22
 ||||| ||||| ||||| |||||
 Db 539 TGCCAGGCAGAGCCAGGG 519

RESULT 12
 AAV00641/c
 ID AAV00641 standard; cDNA: 2816 BP.

XX AC AAV00641;

XX DT 27-MAR-1998 (first entry)

XX DE cDNA encoding human endothelial PAS domain protein 1 (EPAS1).

XX KW Endothelial PAS domain protein 1; EPAS1; screening assay; gene therapy;
 KW endothelial tissue specification; EPAS1 binding; heat shock protein;
 KW basic helix-loop-helix motif; bHLH; hypoxia inducible factor;
 KW HIF-1 alpha binding site; binding affinity; genetic hybridisation screen;
 KW immunogen; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 150...2760
 XX FT /*tag= a

XX PN US5695963-A.

XX PD 09-DEC-1997.

XX PF 17-JAN-1997; 97US-0785241.

XX PR 17-JAN-1997; 97US-0785241.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI McKnight SL, Russell DW, Tian H;

XX DR WPI: 1998-041300/04.

XX DR P-PSDB; AAW37097.

XX PT Endothelial PAS domain protein 1 proteins - used for isolating EPAS1
 XX regulators, e.g. a heat shock protein

XX PS Claim 9; Columns 19-22; 23pp; English.

XX CC The present sequence encodes a novel human endothelial PAS domain
 CC protein 1 (EPAS1). EPAS1 proteins can regulate specification of
 CC endothelial tissue, such as vasculature, the blood-brain barrier, etc..
 CC The protein can be used in a screening assay for agents that
 CC modulate binding of EPAS1 to a binding target, especially a basic

CC helix-loop-helix (bHLH)/PAS protein, a heat shock protein or hypoxia
 CC inducible factor (HIF-1 alpha) binding site. The EPAS1 protein is
 CC incubated with the binding target and a test agent and the effect of the
 CC test agent on the binding affinity of the protein for the target is
 CC determined. The proteins may be produced recombinantly from transformed
 CC host cells from the subject EPAS1 encoding nucleic acids or purified from
 CC mammalian cells. The proteins may be used in diagnosis (e.g. genetic
 CC hybridisation screens for EPAS1 transcripts), therapy (e.g. gene therapy
 CC to modulate EPAS1 gene expression) and in the biopharmaceutical
 CC industry (e.g. as immunogens, reagents for isolating B-cell specific
 CC activators or other transcriptional regulators, reagents for screening
 CC chemical libraries for lead pharmacological agents, etc.).

XX SQ Sequence 2816 BP; 669 A; 880 C; 764 G; 503 T; 0 other;

Query Match 77.4%; Score 17.8; DB 19; Length 2816;
 Best Local Similarity 90.5%; Pred. No. 2.1e-02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagggccgaagcctgg 21
 ||||| ||||| ||||| |||||
 Db 2784 CTGCCAGGTTAGAGGCTGG 2764

RESULT 13
 AAZ94053/c

ID AAZ94053 standard; cDNA: 2818 BP.

XX AC AAZ94053;

XX DT 05-JUN-2000 (first entry)

XX DE DNA encoding human endothelial PAS domain protein-1.

XX KW Endothelial PAS domain protein-1; EPAS1; human; angiogenesis;
 KW antiarteriosclerotic; antitumour; atherosclerosis; tumour;
 KW vascular disease; vulnary; cardiant; vasotropic;
 KW cerebroprotective; gene therapy; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 150...2762
 XX FT /*tag= a

XX PN WO200009657-A2.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18539.

XX PR 14-AUG-1998; 98US-0096515.

XX PA (HARD) HARVARD COLLEGE.

XX PI Lee M, Maemura K, Hiesh C;

XX DR WPI: 2000-205996/18.

XX DR P-PSDB; AAY79161.

XX PT Modulation of angiogenesis in mammals, useful for treating e.g.
 XX atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue
 XX damage, ischemia, balloon angioplasty, frostbite, gangrene or poor
 XX circulation

XX PS Disclosure; Page 14-15; 57pp; English.

XX CC The present sequence is that of cDNA encoding human endothelial PAS
 CC domain protein-1 (EPAS1, see AAY79161). The invention is based on
 CC the discovery that EPAS1 binds to cis-acting regulatory sequences
 CC associated with genes encoding angiogenic factors such as vascular
 CC endothelial cell growth factor (VEGF) and VEGF receptors such as

CC KDR/flk-1 and flt-1, thereby transactivating the promoters of such
CC genes. A claimed method of inhibiting angiogenesis in a mammal
CC comprises administering to the mammal a compound which inhibits
CC binding of EPAS1 to the cis-acting transcription regulatory DNA of
CC an angiogenic factor (see AA294051). The compound may be an EPAS1
CC polypeptide lacking a transactivation domain (see AA79160) or a
CC nucleic acid encoding such a polypeptide, an antisense nucleic acid
CC complementary to mRNA encoding EPAS1, or an EPAS1-specific antibody.
CC The compound is preferably administered to an atherosclerotic
CC lesion or to a tumour site. Angiogenesis is also inhibited using a
CC compound, such as an EPAS1 dominant negative mutant, which inhibits
CC binding of EPAS1 to the EPAS1-binding element, ABNT4 (see AA79162).
CC Angiogenesis can be promoted by administering EPAS1 DNA to increase
CC expression of VEGF or VEGF receptor in endothelial cells of a
CC patient suffering from peripheral vascular disease, cerebral
CC vascular disease, hypoxic tissue damage (e.g. hypoxic damage to
CC heart tissue), or coronary vascular disease as well as to treat
CC patients who have, or have had, transient ischaemic attacks,
CC vascular graft surgery, balloon angioplasty, frostbite, gangrene,
CC or poor circulation.
XX
SQ Sequence 2818 BP; 669 A; 882 C; 764 G; 503 T; 0 other;

Query Match 77.4%; Score 17.8; DB 21; Length 2818;
Best Local Similarity 90.5%; Pred. NO. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgcccagcgaagcctgg 21
||||||| |||||
DB 2784 CTGCCAGGTAGAGGCTGG 2754

RESULT 14
AAZ00365/C
ID AAZ00365 standard; DNA; 3217 BP.

XX AC AAZ00365;

XX DT 26-OCT-1999 (first entry)

XX DE Nucleotide sequence of murine mmFATP1.

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
XX Mus sp.

XX OS

XX PN WO9936537-A2.

XX PD 22-JUL-1999.

XX PF 14-JAN-1999; 99WO-US00182.

XX PR 14-JAN-1999; 99US-0232201.

XX PR 15-JAN-1998; 98US-0071374.

XX PR 20-JUL-1998; 98US-0093491.

XX PR 04-DEC-1998; 98US-0110941.

XX PR 14-JAN-1999; 99US-0232195.

XX PR 14-JAN-1999; 99US-0232197.

XX PR 14-JAN-1999; 99US-0232200.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX PF Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

XX WPI; 1999-444398/37.

XX P-PSDB; AAY14955.

XX Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
XX

PS Example 1; Fig 62; 255pp; English.
XX The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 3217 BP; 616 A; 932 C; 911 G; 758 T; 0 other;

Query Match 77.4%; Score 17.8; DB 20; Length 3217;
Best Local Similarity 90.5%; Pred. NO. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tgcccagcgaagcctgg 22
||||||| |||||
DB 517 TGCCAGGCAGAGGCCAGGG 497

RESULT 15

AAZ00356/C
ID AAZ00356 standard; DNA; 3694 BP.

XX AC AAZ00356;

XX DT 26-OCT-1999 (first entry)

XX DE Nucleotide sequence of human hsfATP1.

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
XX Homo sapiens.

XX OS

XX PN WO9936537-A2.

XX PD 22-JUL-1999.

XX PF 14-JAN-1999; 99WO-US00182.

XX PR 14-JAN-1999; 99US-0232201.

XX PR 15-JAN-1998; 98US-0071374.

XX PR 20-JUL-1998; 98US-0093491.

XX PR 04-DEC-1998; 98US-0110941.

XX PR 14-JAN-1999; 99US-0232195.

XX PR 14-JAN-1999; 99US-0232197.

XX PR 14-JAN-1999; 99US-0232200.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX PF Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

XX WPI; 1999-444398/37.

XX P-PSDB; AAY14946.

XX Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
XX Claim 31; Fig 44A-C; 255pp; English.

XX

The invention provides a family of fatty acid transport proteins (FATPs) that mediate transport of long chain fatty acids (LCFAs) across cell membranes into cells. Human and murine FATP proteins and nucleic acids encoding the proteins are provided. The FATP proteins can be produced by standard recombinant methodology. Fatty acid uptake by cells can be modulated by modulating biosynthesis of FATP proteins especially FATP6. In particular, antisense oligonucleotides can be used to modulate FATP biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid uptake in cardiac muscle of humans. Agents can be directed to cardiac muscle or liver by administration of a complex of the agent and a FATP6 binding moiety. DNA encoding FATP proteins can be used as a reference used in detecting variant alleles or homologues. Altering the LCFA uptake by administering an inhibitor or enhancer of FATP transport function in the small intestine can decrease or increase calories available as fats, and can decrease or increase circulating fatty acids. Blocking the function of FATP4 and also FATP2, is useful for treating obesity, diabetes and heart disease.

XX

SQ Sequence 3694 BP; 622 A; 1202 C; 1110 G; 760 T; 0 other;

Query Match

Best Local Similarity 77.4%; Score 17.8; DB 20; Length 3694;

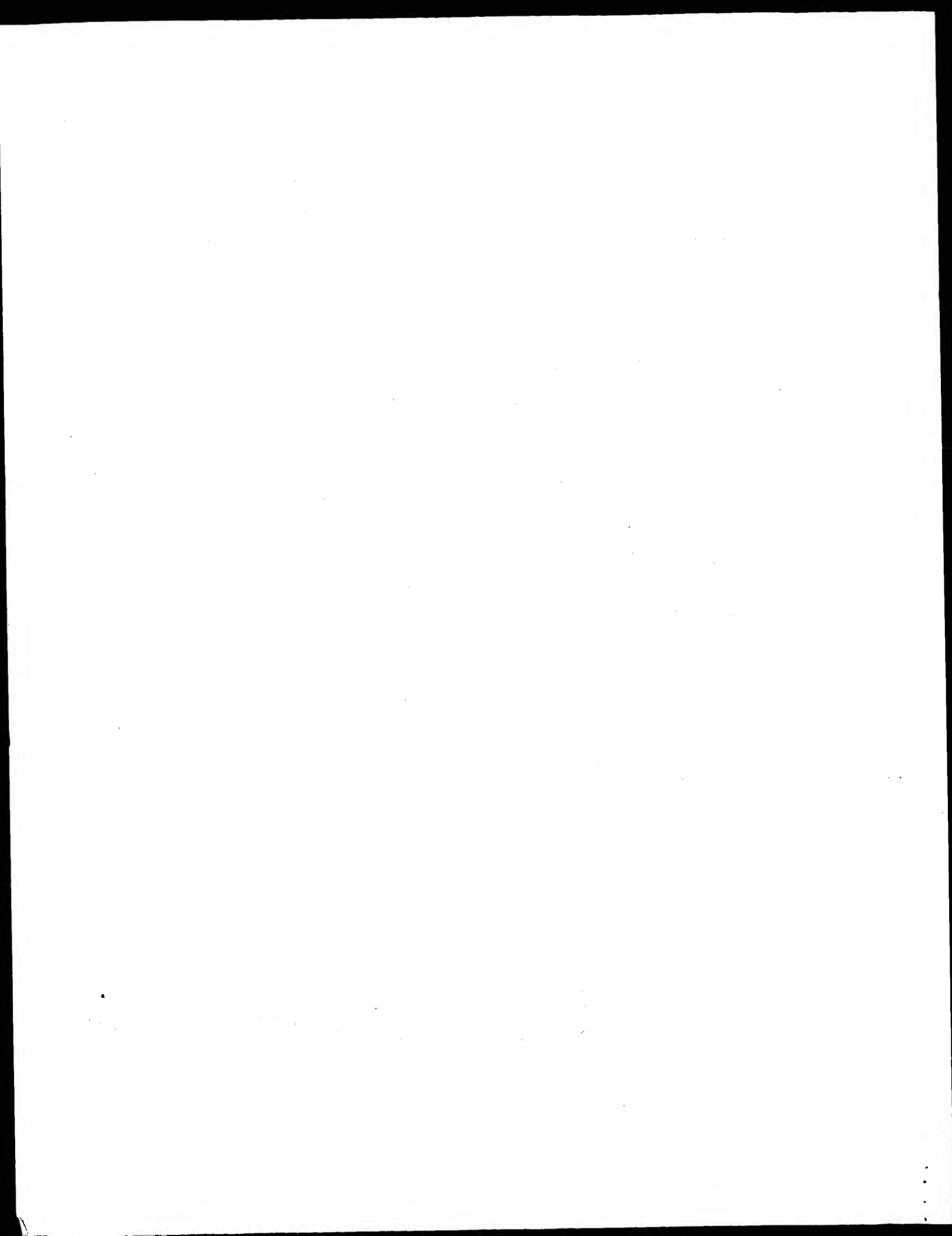
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 tgcccagggccgaagcctggg 22

||||||| ||||||| |||

Db 691 TGCCCAAGGCAGAGGCGAGGG 671

Search completed: October 24, 2001, 10:07:30
Job time: 635 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 13:13:59 ; Search time 5479.82 Seconds
(without alignments)
39.676 Million cell updates/sec

Title: US-09-462-955-6

Perfect score: 23

Sequence: 1 ctgccagggccgaagcctggga 23

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_estl1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
C 1	19.8	86.1	935	152	BG327577		BG327577 602426566
C 2	19.8	86.1	959	144	BF119646		BF119646 601757514
C 3	19.4	84.3	405	117	AW522130	UI-R-BJOP	AW522130 UI-R-BJOP
C 4	19.4	84.3	512	19	AI406571		AI406571 EST234857
C 5	19.4	84.3	517	148	BF403937		BF403937 UI-R-CAL-
C 6	18.8	81.7	176	141	BE860673	UI-M-AHL-	BE860673 UI-M-AHL-
C 7	18.8	81.7	238	126	BE147701	BB147701	BE147701 BB147701
C 8	18.8	81.7	313	114	AW336731		AW336731 21454 MAR
C 9	18.8	81.7	393	130	BE284196		BE284196 BB284196
C 10	18.8	81.7	462	169	BF775369		BF775369 285513 MA
C 11	18.4	80.0	438	3	AA158436	zo59n11.r	AA158436 zo59n11.r
C 12	18.4	80.0	455	117	AW553523		AW553523 L0228C11-
C 13	18.2	79.1	286	161	BB573895		BB573895 BB573895
C 14	18.2	79.1	317	139	BE757761		BE757761 212291 MA
C 15	18.2	79.1	318	241	AZ287293	RPCI-23-1	AZ287293 RPCI-23-1
C 16	18.2	79.1	377	147	BF370922		BF370922 MR0-FN009
C 17	18.2	79.1	392	21	AI520113		AI520113 LD40089.5
C 18	18.2	79.1	453	172	BF993725		BF993725 IL5-GN017
C 19	18.2	79.1	461	115	AW366574		AW366574 RC2-HT014
C 20	18.2	79.1	464	104	AI949053	wg36f03.x	AI949053 wg36f03.x
C 21	18.2	79.1	512	22	AI565947		AI565947 tg68e06.x
C 22	18.2	79.1	542	121	AW80881	QV0-OT003	AW80881 QV0-OT003
C 23	18.2	79.1	556	229	AQ479885	RPCI-11-2	AQ479885 RPCI-11-2
C 24	18.2	79.1	773	168	BF700908		BF700908 602128051
C 25	18.2	79.1	886	151	BF666792		BF666792 602121104
C 26	18.2	79.1	1067	221	CNS03NKE		AL252167 Tetraodon
C 27	17.8	77.4	243	107	AU058457		AU058457 AU058457
C 28	17.8	77.4	295	121	AW877400		AW877400 MR4-PT005
C 29	17.8	77.4	345	168	BF721502		BF721502 mab17h10.
C 30	17.8	77.4	349	167	BE385251		BE385251 601277687
C 31	17.8	77.4	363	168	BF722349	mab17h10.	BF722349 mab17h10.
C 32	17.8	77.4	370	140	BE771160		BE771160 RC5-F007
C 33	17.8	77.4	371	147	BF372847		BF372847 RC5-F007
C 34	17.8	77.4	372	140	BE771169		BE771169 RC5-F007
C 35	17.8	77.4	384	115	AW416678		AW416678 52227 MAR
C 36	17.8	77.4	390	144	BF087382		BF087382 QV2-HT054
C 37	17.8	77.4	397	118	AI326029	mr30e09.x	AI326029 mr30e09.x
C 38	17.8	77.4	397	115	AW408256		AW408256 UI-HF-BK0
C 39	17.8	77.4	401	121	AW382777		AW382777 QV2-LT005
C 40	17.8	77.4	409	141	BE974489		BE974489 601432319
C 41	17.8	77.4	412	19	AI386449		AI386449 mr30e09.y
C 42	17.8	77.4	418	140	BE818183		BE818183 RC6-BN027
C 43	17.8	77.4	430	162	BE014102		BE014102 125686 MA
C 44	17.8	77.4	433	162	BE013607		BE013607 125013 MA

ALIGNMENTS

RESULT	1	Score	Query %	Match	Length	DB	ID	Description
LOCUS	BG327577		935 bp	mrNA	EST		27-FEB-2001	
DEFINITION	602426566F1 NIH_MGC_14		602426566F1 NIH_MGC_14		Homo sapiens cDNA clone IMAGE:4564192 5',			
ACCESSION	BG327577		mrNA sequence.					
VERSION	BG327577.1		GI:13134015					
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.							
TITLE	NIH-MGC http://mgi.nci.nih.gov/							
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)							
COMMENT	Unpublished (1999)							
	Contact: Robert Strausberg, Ph.D.							

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI279 row: d column: 17
High quality sequence stop: 789.

FEATURES

source
Location/Qualifiers
1..935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4564192"
/clone_lib="NIH_MGC_14"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(3). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

222 a 236 c 297 g 180 t

Query Match

Rest Local Similarity 86.1%; Score 19.8; DB 152; Length 935;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagccgaagcctggga 23

||||||| |||||||

Db 910 CTGCCAGCGGAGGCGCTGGCA 888

RESULT 2

BF119646/c

DEFINITION

LOCUS

BF119646

BF119646

BF119646.1 GI:10958790

EST.

house mouse.

Mus musculus

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Lotmar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI279 row: 1 column: 03

High quality sequence stop: 612.

Location/Qualifiers

1..959

/organism="Mus musculus"

/strain="C57/B6"

/db_xref="taxon:10090"

/clone="IMAGE:3986834"

/clone_lib="NCI_CGAP_Mam5"

ACCESSION BF403937
 VERSION BF403937.1 GI:11391912
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 517)
 REFERENCE Bernaldo, M.F., Lennon, G. and Soares, M.B.
 AUTHORS Normalization and subtraction: two approaches to facilitate gene
 TITLE discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized thalamus library cDNA Library preparation: M.B. Soares
 Lab Clone distribution: clones will be available through Research
 Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1..517

/organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_lib="UI-R-CAI-bid-j-19-0-UI"
 /clone_lib="UI-R-CAI"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CAI
 library is a subtracted library derived from the following
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons
 , midbrain, cerebral cortex, corpus striatum, testis, and
 hippocampus. For a detailed description of the library
 from which this clone was derived, please visit our web
 site at ratest.eng.uiowa.edu. The subtraction has been
 previously described in (Bernaldo, Lennon and Soares,
 Genome Research 6:791-806, 1996)
 TAG_LIB=UI-R-CAI
 TAG_TISSUE=thalamus
 TAG_SEQ=ATCGC"

BASE COUNT 114 a 134 c 133 g 136 t
 ORIGIN

Query Match 84.3%; Score 19.4; DB 148; Length 517;
 Best Local Similarity 95.2%; Pred. No. 3.9e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gccacggccgaagcctggga 23
 ||| ||||| ||||| |||||

Db 313 GCCGAGCGCGAAGCGCTGGG 293

RESULT

6

BB860673
 LOCUS UI-M-AH1-agv-c-04-0-UI.r1 NIH_BMAP_MCE.N Mus musculus cDNA clone
 DEFINITION 29-SEP-2000
 EST
 UI-M-AH1-agv-c-04-0-UI 5', mRNA sequence.
 ACCESSION BE860673
 VERSION BE860673.1 GI:10377856
 KEYWORDS EST.

SOURCE

house mouse.
 Mus musculus

REFERENCE Bernaldo, M.F., Lennon, G. and Soares, M.B.
 AUTHORS Normalization and subtraction: two approaches to facilitate gene
 TITLE discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mEST@mail.nih.gov
 cDNA Library Preparation: M.B. Soares Lab Clone distribution:
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
 should be noted that Bento Soares is generating a small number of
 additional specialized non-redundant arrays of BMAP cDNAs whose
 availability will be considered under appropriate and limited
 collaborative arrangements

FEATURES

Location/Qualifiers
 1..176

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="UI-M-AH1-agv-c-04-0-UI"
 /clone_lib="NIH_BMAP_MCE.N"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NIH_BMAP_MCE library is a normalized library constructed
 from mouse cerebellum. The tag is a string of 5
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bernaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996. Tissue provided by Ms. Annie Novakovich,
 Zivic-Miller Laboratories."

BASE COUNT 29 a 58 c 40 g 49 t
 ORIGIN

Query Match 81.7%; Score 18.8; DB 141; Length 176;
 Best Local Similarity 90.9%; Pred. No. 6.7e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccacggccgaagcctggg 22
 ||| ||||| ||||| |||||

Db 71 CTGCTCAGCCCGAAGCGCTGGG 92

RESULT

7

BB147701
 LOCUS BB147701 RIKEN full-length enriched, adult female vagina Mus
 DEFINITION musculus cDNA clone 9930101123 3', mRNA sequence.
 ACCESSION BB147701
 VERSION BB147701.1 GI:8802638
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 238)
 REFERENCE Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
 , P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.


```

plate: I0228 row: C column: 11
Seq primer: -21M13 Forward
High quality sequence stop: 455
POLYA=Yes.
Location/Qualifiers
1. .455
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L0228C11"
/clone.lib="Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/note="vector: pSPORT1 (Gibco/BRL Life Technology);
Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an Oligo(dT) primer [NotI
primer-adaptor from Gibco/BRL]
[5',_PGAGTAGTCTAGATCGAGCGCCGCTTTT-3'] from
2.5ug of total RNA. The double-stranded cDNAs were
treated with 14 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
lone-linker L1-Sal3 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from
free linkers by Centricon 100. Then, cDNAs were amplified
by long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's size fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Xiaohong Wang and Yulan Piao."

```

BASE COUNT	109 a	152 c	158 g	36 t
ORIGIN				

```
Query Match      80.0%; Score 18.4; DB 117; Length 455;
Best Local Smilarity 95.0%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 3 gccagggccgaaggcctggg 22
||||| |||||||
Db 265 GCCCAGGCTGAAGGCCCTGGG 284

[illegible]

MUSCULUS CDNA CLONE 45340004 5' (partial sequence)

 ACCESSION BB573895
 VERSION BB573895.1 GI:11470439
 KEYWORDS EST.
 SOURCE house mouse.

ORGANISM	Mus musculus
REFERENCE	1 (bases 1 to 286)
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AIZAWA, K., AKAHIRA, S., AKIMURA, T., ARAI, A., ARAKAWA, T., CARNINCI, L., HANAGAKI, T., HAYATO, S., HIROKATA, T., HIROZANE, T., HODOYAMA, Y., IMOTANI, K., ISHII, Y., ITOH, M., IZAWA, M., KAWAI, J., KOJIMA, Y., KONNO, H., KUSAKABE, M., MATSUYAMA, T., MIYAZAKI, A., NAKAMURA, M., NISHI, K., NOMURA, K., NUMAZAKI, R., OKAZAKI, Y., OKIDO, T., OWA, C., SAKAI, C., SAKAI, K., SAKAI, D., SATO, K., SHIBATA, K., SHIBATA, Y., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y., SUZUKI, H., TAGAWA, A., TAKAKASHI, F., TANAKA T., TOYA, T., WATAHAKI, A., YAMAMURA, T., YASUNISHI, A., YOSHIDA, K., YOSHIOKI, A., MURAMATSU, M. and HAYASHIZAKI, Y.

TITLE	RIKEN MOUSE ESTs (Aizawa, K. et al. 2000)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshihide Hayashizaki

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 438)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
Jr., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
White, X., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson, RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 2561 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 392.

```

FEATURES             Location/Qualifiers
     i. .438
    /organism="Homo sapiens"
    /db_xref="GDB:4622623"
    /db_xref="taxon:9606"
    /clone="IMAGE:591237"
    /clone_lib="Stratagene pancreas (#937208)"
    /lab_host="SOuR cells (kanamycin resistant)"
    /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTGTGTGTGTGTGTGTGT 3'"

```

BASE COUNT	98 a	150 c	115 g	67 t	8 others
ORIGIN					
Query Match			80.0 %;	Score 18.4;	DB 3; Length 438;

Qy	1	ctgcccaggccgaagcctgg	21
Db	263	cnccccaggccgaagacctgg	243

Best local similarity 90.5%; Pred No. le+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT	12
AW553523	
LOCUS	455 bp mRNA EST 31-AUG-2000
DEFINITION	L0228C11-3 Mouse Newborn Ovary cDNA Library Mus musculus cdna clone
	L0228C11 3', mRNA sequence.
ACCESSION	AW553523
VERSION	AW553523.1 GI:7198946
KEYWORDS	EST.

SOURCE	ORGANISM
house mouse.	<i>Mus musculus</i>
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1 (bases 1 to 455)
AUTHORS	Tanaka, T. S., Jaradat, S. A., Lim, M. K., Kargul, G. J., Wang, X., Grahovac, D. A., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W. H., M. J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W. H., I. J., Becker, K. G. and Ko, M. S. H.
TITLE	Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE	20381348
COMMENT	Contact: George J. Kargul

Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov

Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermostabilization of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawaj, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (3), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

Location/Qualifiers
1. .286

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="493340M04"
/clone_lib="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT

ORIGIN 62 a 74 c 94 g 56 t

Query Match

Best Local Similarity 79.1%; Score 18.2; DB 161; Length 286;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ctgcccaggccgaagcctggga 23
||||| ||| ||||| ||||| |||||
Db 157 CTGCCCGGGCGGAGAGCCTGGGA 179

RESULT 14

LOCUS BE757761/c 317 bp mRNA EST 15-SEP-2000
DEFINITION 212291 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757761
VERSION BE757761.1 GI:10171753
KEYWORDS EST.
SOURCE cOW.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE

1 (bases 1 to 317)

AUTHORS

Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,
Bennett, G.A., Fahrenkrug, S.C., Fekking, B.A., Rohrer, G.A., Laegreid,
W.W. and Keele, J.W.

TITLE

Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle

JOURNAL

Unpublished (2000)

COMMENT

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers

FORWARD: AGCAACACAGCTATGACCAT

BACKWARD: GTTTCCAGTCACGAGC

Plate: 65 row: N column: 2

Seq primer: ATTAGTGCACACTATAG.

FEATURES

source

1. .317

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 2BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;

Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

BASE COUNT 64 a 60 c 77 g 116 t

Query Match

Best Local Similarity 79.1%; Score 18.2; DB 139; Length 317;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY

1 ctgcccaggccgaagcctggga 23

Db

87 CTGCTCAGGCTCAGGCTGGGA 65

RESULT 15

AZ287293

LOCUS

AZ287293

DEFINITION

RPCI-23-144118.TJ RPCI-23 Mus musculus genomic clone RPCI-23-144118

VERSION

AZ287293

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 318)

AUTHORS

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret,
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.

TITLE

Mouse BAC End Sequences from Library RPCI-23

JOURNAL

Unpublished (1999)

COMMENT

Other_GSSs: RPCI-23-144118.TV

Contact:

Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Thu Oct 25 13:09:06 2001

Pax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 144 row: 1 column: 18

Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1. .318
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-144118"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
 56 c 99 g 74 t

BASE COUNT
 ORIGIN

Query Match 79.1%; Score 18.2; DB 241; Length 318;
 Best Local Similarity 87.0%; Pred. No. 1.2e+03;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ctgcccaggccgagcctggga 23
 Db 92 CTGCCAGGAGAGACCTGAGA 114

Search completed: October 24, 2001, 13:14:03
 Job time: 11828 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 11:42:58 ; Search time 5701.85 Seconds
(without alignments)
62.393 Million cell updates/sec

Title: US-09-462-955-6
Perfect score: 23

Sequence: 1 ctgccaggccgaaggctggga 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_bal.*
2: gb_bal.*
3: gb_bal.*
4: gb_bal.*
5: gb_bal.*
6: gb_bal.*
7: gb_bal.*
8: gb_bal.*
9: gb_bal.*
10: gb_bal.*
11: gb_bal.*
12: gb_bal.*
13: gb_bal.*
14: gb_bal.*
15: gb_bal.*
16: gb_bal.*
17: gb_bal.*
18: gb_bal.*
19: gb_bal.*
20: gb_bal.*
21: gb_bal.*
22: gb_bal.*
23: gb_bal.*
24: gb_bal.*
25: gb_bal.*
26: gb_bal.*
27: gb_bal.*
28: gb_bal.*
29: gb_bal.*
30: gb_bal.*
31: gb_bal.*
32: gb_bal.*
33: gb_bal.*
34: gb_bal.*
35: gb_bal.*
36: gb_bal.*
37: gb_bal.*
38: gb_bal.*
39: gb_bal.*
40: gb_bal.*
41: gb_bal.*
42: gb_bal.*
43: gb_bal.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vl.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vil.*
59: gb_vil.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_rol.*
95: gb_rod.*
96: gb_in4.*
97: gb_pr10.*
98: em_bal.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	1291	58	CFDCG
c	19.8	86.1	25010	90	M29963 Coconut fol
3	19.8	86.1	185002	79	AL353092 Human DNA
c	19.8	86.1	193168	62	AL161784 Homo sapi
5	19.8	86.1	204881	81	AC011969 Homo sapi
c	19.8	86.1	207392	60	AL513122 Homo sapi
7	19	82.6	15982	1	AC006558 Homo sapi
8	19	82.6	169194	74	AE004779 Pseudomon
					AC073464 Homo sapi

```

c 9 19 82.6 170237 89 AL137070 Human DNA
10 18.8 81.7 5162 2 AF151965 Pseudomon
c 11 18.8 81.7 8660 3 AF153207 Pseudomon
c 12 18.8 81.7 14750 3 SCF1 AL117322 Streptomy
c 13 18.8 81.7 176253 82 AP001078 Homo sapi
c 14 18.8 81.7 193487 82 AP001793 Homo sapi
c 15 18.8 81.7 210734 63 AC015676 Homo sapi
c 16 18.8 81.7 212055 82 AP000899 Homo sapi
c 17 18.4 80.0 4091 2 AF323675 Bradyrhiz
c 18 18.2 79.1 189 95 RATPA08 M31191 Rat tissue-
c 19 18.2 79.1 1103 97 HSU13668 U13668 Human G pro
c 20 18.2 79.1 1262 97 HUMGPCRD L32831 Homo sapien
c 21 18.2 79.1 1982 93 HSACCCAGEN X83956 H sapiens A
c 22 18.2 79.1 2022 89 AK026727 Homo sapi
c 23 18.2 79.1 2445 95 RATPATISS M23697 Rat tissue-
c 24 18.2 79.1 2512 9 A19618 A19618 Mammalian c
c 25 18.2 79.1 2526 93 HSCBEP A Y11525 H sapiens m
c 26 18.2 79.1 3542 97 HSU18550 U18550 Human GPR3
c 27 18.2 79.1 4208 92 HSA010482 AJ010482 Homo sapi
c 28 18.2 79.1 11263 97 HSU67167 U67167 Homo sapien
c 29 18.2 79.1 12237 97 HSE68061 U68061 Human MUC2
c 30 18.2 79.1 43487 6 CELC13D9 AF016420 Caenorhab
c 31 18.2 79.1 63609 89 AL133343 Human DNA
c 32 18.2 79.1 77322 93 HSDJ144C9 AL096774 Human DNA
c 33 18.2 79.1 121417 80 AL359374 AC011651 Homo sapi
c 34 18.2 79.1 125111 62 AC011651 AC0178603 Human DNA
c 35 18.2 79.1 150355 33 HSJ364H10 AC021287 Homo sapi
c 36 18.2 79.1 154385 66 AC021287 AC011043 Homo sapi
c 37 18.2 79.1 154701 80 AC011043 AC041043 Homo sapi
c 38 18.2 79.1 155067 71 AC041043 AC024535 Homo sapi
c 39 18.2 79.1 158463 68 AC024535 AC0358236 Homo sapi
c 40 18.2 79.1 164842 80 AC0358236 AC011935 Homo sapi
c 41 18.2 79.1 168043 62 AC011935 AC0590022 Homo sapi
c 42 18.2 79.1 171049 82 AC0590022 AC016992 Homo sapi
c 43 18.2 79.1 186765 64 AC016992 AC023331 Homo sapi
c 44 18.2 79.1 192180 67 AC023331 AC016725 Homo sapi
c 45 18.2 79.1 203356 64 AC016725 Homo sapi

```

ALIGNMENTS

```

RESULT 1
CFDCG 1291 bp ss-DNA circular VRL 02-AUG-1993
LOCUS Coconut foliar decay virus, complete genome.
DEFINITION M29963
ACCESSION circular; complete genome.
VERSION Coconut foliar decay virus DNA, clones NMORG, pORG, and pORGSAU3A.
KEYWORDS Coconut foliar decay virus
SOURCE Viruses; ssDNA viruses; Nanovirus.
ORGANISM
REFERENCE 1 (bases 1 to 1291)
AUTHORS Rohde, W., Randles, J.W., Langridge, P. and Hanold, D.
TITLE Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus
JOURNAL Virology 176, 648-651 (1990)
MEDLINE 90266484
COMMENT Draft entry and printed sequence for [1] kindly submitted by W.Rhode, 15-MAR-1989, for release after publication.
FEATURES
source
Location/Qualifiers
1..1291
/organism="Coconut foliar decay virus"
/db_xref="taxon:12474"
40..70
/note="stem-loop structure"
103..975
/note="ORF 1"
/codon_start=1
/protein_id="AAA42894.1"
/db_xref="GI:323307"
/translation="MGSSTRMWCFTLVNTEFEFAANVVRRIEFLNVAIVGDEVAPS
TGQRHLQGGFIHLKTRRUGLOGLKTVLGNDRHLEPTRGSDQNRDVCSEKRVLLHEGVP

```

```

TRPGVKRPLAQRFAEEFDELRLDPGGRVRCVVGASVETRWAAENPFFPHNKO
LEVSATGEPADRTILWICRGDGGKSVFAYLGLKPDWFTCGTRKDVLYQYLE
DPKRNLLILDPRCNLEYLNVALLECVKNRAFSDDKYEPILGFDHVLVFNVLDP
YLKTSRDRILKLN1"
314..775
/note="ORF 2"
/codon_start=1
/protein_id="AAA42895.1"
/db_xref="GI:323308"
/translation="MTGFTWSPVVPVPTNRIETTVRRNGCFSSSTESRLVLSKSHDWPN
DLNRKLNMSANKTOADTEDALYTELRWNGDGPGLKIRSHFHTITIGSLKCLCRSEORT
IAQSSGYADETEEGSPCLNLSSTSPGSHVVEPERVYCTSTSRITQNEI"
complement(422..568)
/note="ORF 6"
/codon_start=1
/protein_id="AAA42896.1"
/db_xref="GI:323309"
/translation="MEMGTDFQRILSVIPKLRVQIRGLGLPGGVHQPQIVGP
IVAF"
639..797
/note="ORF 3"
/codon_start=1
/protein_id="AAA42897.1"
/db_xref="GI:323310"
/translation="MRTRRRREVRCQISRTQARLVLLHMMWNQKGRIVPVHGRPKTK
FNPRCTQV"
complement(823..987)
/note="ORF 5"
/codon_start=1
/protein_id="AAA42898.1"
/db_xref="GI:323311"
/translation="WHTLNLIPQFYSVPADFOIIRQDIDGKYEYMHMVEPKITKGFVVF
RTECPVLN1TF"
1098..1286
/note="ORF 4"
/codon_start=1
/protein_id="AAA42899.1"
/db_xref="GI:323312"
/translation="MNRVMGGPTIKDSIWRTNLLCLCLOCTPLSTSIQVSSLLLEKK
AASLYLPSICFCAIGRLS"
BASE COUNT 336 a 323 c 332 g 300 t
ORIGIN
Query Match 100.0%; Score 23; DB 58; Length 1291;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ctgccagccgaagcctggga 23
|||||
Db 1045 CTGCCAGGCCGGAAGCCTGGGA 1067

```

RESULT 2

```

AL353092/c 25010 bp DNA PRI 11-FEB-2001
LOCUS Human DNA sequence from clone RP1-180113 on chromosome 20 Contains
DEFINITION 5' end of the HCK gene for hemopoietic cell kinase (protein
tyrosine kinase), contains ESTs, STSS, GSSs and a CpG island,
complete sequence.
ACCESSION AL353092
VERSION AL353092.6 GI:9650539
KEYWORDS HTG; CpG island; HCK; tyrosine kinase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 25010)
AUTHORS Almeida, J.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

```

COMMENT

On Aug 1, 2000 this sequence version replaced gi:9187765.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlap in the sequence submission
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep/ This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
 IMPORTANT: This sequence is not the entire insert of clone
 RP1-180113 It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP5-836N17 is at 24911 in this sequence.
 The true right end of clone RP1-310013 is at 100 in this sequence.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP1-180113 is from
 the library RPC1-1 constructed by the group of Pieter de Jong. For
 further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2.

FEATURES

Source	Location/Qualifiers
	1..25010
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="20"
	/clone="RP1-180113"
	/clone_lib="RPC1-1"
repeat_region	1..43
	/note="AluJo/FLAM repeat: matches 76..118 of consensus"
misc_feature	complement(39..567)
	/note="match: GSS: Em:AQ386884"
repeat_region	58..175
	/note="5S repeat: matches 1..119 of consensus"
repeat_region	265..455
	/note="MER20 repeat: matches 26..218 of consensus"
repeat_region	510..586
	/note="L2 repeat: matches 1966..2045 of consensus"
misc_feature	692..906
	/note="match: GSS: Em:AQ486790"
repeat_region	772..843
	/note="2 copies 36 mer 93% conserved"
misc_feature	1576..2453
	/note="CpG island"
	/evidence=not_experimental
mRNA	join(1846..1956,21132..21252,22789..22831,23186..23288,24093..>24191)
	/gene="HCK"
	/note="match: cDNAs: Em:M16591 Em:X62345 Em:S74141 Em:J03023 Em:M16592 Em:M83666 Em:Y00487 Em:X60380 Em:J03579 Em:X67786 Em:AF000300 Em:AF000301 Em:AF000302 Em:M17031 Em:X52822 Em:X57018 Em:X15345 Em:M27454 Em:M57290 Em:X57684 Em:M85043 Em:M19722 match: ESTs: Em:AW307786"
	/evidence=not_experimental
	/product="dJ180113.1 (hemopoietic cell kinase)"
gene	1846..24191
	/gene="HCK"
misc_feature	complement(2102..2519)
	/note="match: GSS: Em:AQ382576"
misc_feature	complement(2108..2557)
	/note="match: GSS: Em:AQ461389"
	complement(2196..2570)
	/note="match: GSS: Em:AQ240986"
repeat_region	2546..2600
	/note="MIR repeat: matches 89..148 of consensus"
misc_feature	2575..3113
	/gene="HCK"
	/note="match: GSS: Em:AQ201482"
misc_feature	2576..3351
	/gene="HCK"
	/note="match: GSS: Em:AQ748323"
misc_feature	2594..2974
	/gene="HCK"
	/note="match: GSS: Em:AQ821722"
repeat_region	2700..2825
	/note="MIR repeat: matches 76..193 of consensus"
repeat_region	2826..3131
	/note="AluSx repeat: matches 1..306 of consensus"
repeat_region	3132..3189
	/note="MIR repeat: matches 193..252 of consensus"
repeat_region	3221..3389
	/note="MERSA repeat: matches 1..184 of consensus"
repeat_region	3809..3976
	/note="AluSp repeat: matches 147..313 of consensus"
repeat_region	3997..4296
	/note="AluY repeat: matches 3..302 of consensus"
repeat_region	4866..5115
	/note="L1MC1 repeat: matches 5725..5980 of consensus"
repeat_region	5124..5701
	/note="L1MB5 repeat: matches 5615..5671 of consensus"
repeat_region	5702..5999
	/note="AluY repeat: matches 1..295 of consensus"
repeat_region	6000..6157
	/note="L1MB5 repeat: matches 5454..5615 of consensus"
repeat_region	6276..6378
	/note="L2 repeat: matches 2623..2730 of consensus"
repeat_region	6789..7086
	/note="AluJb repeat: matches 1..299 of consensus"
repeat_region	7087..7147
	/note="MIR repeat: matches 68..133 of consensus"
repeat_region	7614..7706
	/note="MIR repeat: matches 47..140 of consensus"
repeat_region	7720..8014
	/note="AluSq repeat: matches 1..297 of consensus"
misc_feature	7895..8073
	/gene="HCK"
	/note="match: STS: Em:HS336XG5"
misc_feature	8076..8206
	/gene="HCK"
	/note="match: STS: Em:HS336XG5"
repeat_region	8076..8119
	/note="22 copies 2 mer ac 97% conserved"
repeat_region	8139..8510
	/note="LTR16C repeat: matches 1..387 of consensus"
repeat_region	8750..8941
	/note="48 copies 4 mer cttc 62% conserved"
repeat_region	8779..8928
	/note="75 copies 2 mer tt 60% conserved"
repeat_region	8795..8974
	/note="5 copies 36 mer 68% conserved"
repeat_region	8820..8957
	/note="46 copies 3 mer cct 66% conserved"
repeat_region	8982..9285
	/note="AluJo repeat: matches 1..304 of consensus"
repeat_region	9680..10012
	/note="AluSx repeat: matches 1..311 of consensus"
repeat_region	10208..10442
	/note="LTR16B repeat: matches 250..459 of consensus"
repeat_region	10443..10751
	/note="AluSg repeat: matches 1..309 of consensus"
repeat_region	10752..10991
	/note="LTR16B repeat: matches 2..250 of consensus"
repeat_region	11000..11406
	/note="L2 repeat: matches 2157..2577 of consensus"

```

misc_feature      complement(11015..11870)
                  /note="match: GSS: Em:AQ896828"
repeat_region     11407..11707
                  /note="AluX repeat: matches 3..303 of consensus"
misc_feature      complement(11415..11855)
                  /note="match: GSS: Em:AQ369535"
misc_feature      complement(11468..11868)
                  /note="match: GSS: Em:AQ135725"
misc_feature      complement(11535..11868)
                  /note="match: GSS: Em:AQ709950"
repeat_region     11708..11734
                  /note="L2 repeat: matches 2132..2157 of consensus"
misc_feature      11890..12417
                  /gene="HCK"
                  /note="match: GSS: Em:AQ583762"
misc_feature      11920..12309
                  /gene="HCK"
                  /note="match: GSS: Em:AQ419051"
repeat_region     12733..13018
                  /note="AluJ repeat: matches 1..292 of consensus"
repeat_region     13463..13759
                  /note="AluX repeat: matches 5..311 of consensus"
repeat_region     13991..14088
                  /note="MIR repeat: matches 47..144 of consensus"
repeat_region     14544..14681
                  /note="MIR repeat: matches 2..144 of consensus"
repeat_region     14682..14801
                  /note="MER5A repeat: matches 66..187 of consensus"
repeat_region     15409..15538
                  /note="MER5A repeat: matches 4..160 of consensus"
repeat_region     15774..16082
                  /note="AluX repeat: matches 1..310 of consensus"
repeat_region     16689..16903
                  /note="L1MB3 repeat: matches 5953..6176 of consensus"
repeat_region     16936..17021
                  /note="13 copies 2 mer tt 92% conserved"

Query Match      86.1%; Score 19.8; DB 90; Length 25010;
Best Local Similarity 91.3%; Pred: No.1.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctgccagccggaagcctggga 23
    ||||| ||||| ||||| ||
Db 21535 CTGCCAGCGCGGAGCCTGAGA 21513

RESULT 3
AL161784
LOCUS
DEFINITION      AL161784 185002 bp DNA HTG 09-MAR-2001
PROGRESS ***, 12 unordered pieces.
ACCESSION      AL161784
VERSION        AL161784.8 GI:13273607
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                1 (bases 1 to 185002)
                Plumb,B.
REFERENCE      Direct Submission
AUTHORS       Submitted (04-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
TITLE         CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
JOURNAL       requests: clonerequest@sanger.ac.uk
COMMENT       On Mar 11, 2001 this sequence version replaced gi:9796857.
                ----- Genome Center
                Center: Sanger Centre
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquery@sanger.ac.uk
                ----- Project Information
                Center project name: BA342H3
                ----- Summary Statistics

```

```

Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 178919 bases at least Q40
Consensus quality: 181170 bases at least Q30
Consensus quality: 182624 bases at least Q20
Insert size: 183902; sum-of-contigs
Quality coverage: 5.27x in Q20 bases; sum-of-contigs Quality
coverage: 5.51x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 22346: contig of 22346 bp in length
* 22347 22446: gap of 100 bp
* 22447 24632: contig of 2186 bp in length
* 24633 24732: gap of 100 bp
* 24733 31381: contig of 6649 bp in length
* 31382 31481: gap of 100 bp
* 31482 36730: contig of 5249 bp in length
* 36731 36830: gap of 100 bp
* 36831 43040: contig of 6210 bp in length
* 43041 43140: gap of 100 bp
* 43141 56310: contig of 13170 bp in length
* 56311 56410: gap of 100 bp
* 56411 58488: contig of 2078 bp in length
* 58489 58588: gap of 100 bp
* 58589 102210: contig of 43622 bp in length
* 102211 102310: gap of 100 bp
* 102311 175449: contig of 73139 bp in length
* 175450 175549: gap of 100 bp
* 175550 177644: contig of 2095 bp in length
* 177645 177744: gap of 100 bp
* 177745 182725: contig of 4981 bp in length
* 182726 182825: gap of 100 bp
* 182826 185002: contig of 2177 bp in length.
FEATURES
    source
        1..185002 Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone_lib="RPC1-11.2"
            1..22346
            /note="assembly_fragment:00892"
            fragment_chain:1
            clone_end:SP6
            vector_side:left
            22447..24632
            /note="assembly_fragment:02256"
            fragment_chain:1
            24733..31381
            /note="assembly_fragment:02541"
            fragment_chain:2
            31482..36730
            /note="assembly_fragment:02286"
            fragment_chain:2
            36831..43040
            /note="assembly_fragment:02017"
            fragment_chain:2
            43141..56310
            /note="assembly_fragment:01247"
            fragment_chain:2
            56411..58488
            /note="assembly_fragment:01774"
            fragment_chain:3
            58589..102210
            /note="assembly_fragment:01774"
            fragment_chain:3
            58589..102210

```



```

/note="assembly_fragment:01103
fragment_chain:3"
misc_feature
102311..175449
/note="assembly_fragment:00558
fragment_chain:3"
misc_feature
175550..177644
/note="assembly_fragment:00056"
177745..182725
/note="assembly_fragment:00871"
182826..185002
/note="assembly_fragment:02179"
53278 a 41233 c 39095 g 50291 t 1105 others
ORIGIN

```

```

Query Match      86.1%; Score 19.8; DB 79; Length 185002;
Best Local Similarity .91.3%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 ctgccagggccgagcctggga 23
|||||
Db 136974 CTGCCAGGCCAAAGGCGTCAA 136996

```

```

RESULT 4
AC011969/c
LOCUS          AC011969 193168 bp DNA HTG 21-APR-2000
DEFINITION    Homo sapiens chromosome 4 clone RP11-520J8 map 4, WORKING DRAFT
SEQUENCE      AC011969
ACCESSION     AC011969
VERSION       AC011969.3 GI:7630868
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE     1 (bases 1 to 193168)
AUTHORS       Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balaghi,J., Barina,N., Beckerly,R., Boguslavsky,L., Bouckhagalter,B.,
Brown,A., Castle,A., Colangelo,R., Collins,S., Collymore,A.,
Cooke,P., Dearlano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGuck,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 21, 2000 this sequence version replaced gi:7008795.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2649
Center clone name: 520-J8
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

FEATURES

```

Source
1..193168
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-520J8"
/clone_lib="RPC1-11 Human Male BAC"
1..1093
/note="assembly_fragment"
1194..2376
/note="assembly_fragment"
2477..5933
/note="assembly_fragment"
6034..9960
/note="assembly_fragment"
10061..16555
/note="assembly_fragment"
16656..22671
/note="assembly_fragment"
22772..31113
/note="assembly_fragment"
31214..41484
/note="assembly_fragment"
41585..50824
/note="assembly_fragment"
50925..69485
/note="assembly_fragment"
69586..99210
/note="assembly_fragment"
99311..130378
/note="assembly_fragment"
130479..193168
/note="assembly_fragment"

```

```

Assembly program: Phrap; version 0.960731
Consensus quality: 126832 bases at least Q40
Consensus quality: 155745 bases at least Q30
Consensus quality: 175228 bases at least Q20
Insert size: 186000; agarose-fp
Quality coverage: 3.2 in Q20 bases; agarose-fp
Quality coverage: 3.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as-soon as it is available and the accession number will
* be preserved.
*
* 1..1093: contig of 1093 bp in length
* 1094..1193: gap of 100 bp
* 1194..2376: contig of 1183 bp in length
* 2377..2476: gap of 100 bp
* 2477..5933: contig of 3457 bp in length
* 5934..6033: gap of 100 bp
* 6034..9960: contig of 3927 bp in length
* 9961..10060: gap of 100 bp
* 10061..16555: contig of 6495 bp in length
* 16556..22671: contig of 6016 bp in length
* 22672..2771: gap of 100 bp
* 2772..31113: contig of 8342 bp in length
* 31114..31213: gap of 100 bp
* 31214..41484: contig of 10271 bp in length
* 41485..50824: contig of 9240 bp in length
* 50825..5924: gap of 100 bp
* 5925..69485: contig of 18561 bp in length
* 69486..99210: contig of 29625 bp in length
* 99211..99310: gap of 100 bp
* 99311..130378: contig of 31068 bp in length
* 130379..193168: gap of 100 bp
* 193169..193168: contig of 62690 bp in length.

```

FEATURES

```

Location/Qualifiers
1..193168
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-520J8"
/clone_lib="RPC1-11 Human Male BAC"
1..1093
/note="assembly_fragment"
1194..2376
/note="assembly_fragment"
2477..5933
/note="assembly_fragment"
6034..9960
/note="assembly_fragment"
10061..16555
/note="assembly_fragment"
16656..22671
/note="assembly_fragment"
22772..31113
/note="assembly_fragment"
31214..41484
/note="assembly_fragment"
41585..50824
/note="assembly_fragment"
50925..69485
/note="assembly_fragment"
69586..99210
/note="assembly_fragment"
99311..130378
/note="assembly_fragment"

```

```

/clone="assembly_fragment
vector_side:left"
misc_feature
130479..193168
/note="assembly_fragment"
BASE COUNT 55031 a 40516 c 42876 g 53468 t 1277 others
ORIGIN

Query Match 86.1%; Score 19.8; DB 62; Length 193168;
Best Local Similarity 91.3%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ctgccagccggaagcctggga 23
|||||
Db 115095 CTGCCAGCCCAAGGCTGGAA 115073
|||||

RESULT 5
AL513122 204881 bp DNA HTG 25-FEB-2001
LOCUS Homo sapiens chromosome 9 clone RP11-477J21, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.
ACCESSION AL513122
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 204881)
Direct Submission
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Feb 28, 2001 this sequence version replaced gi:12578283.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA477J21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 201478 bases at least Q40
Consensus quality: 202815 bases at least Q30
Consensus quality: 203621 bases at least Q20
Insert size: 204181; sum-of-contigs
Insert size: 206137; 6.5% error; agarose-fp
Quality coverage: 6.47x in Q20 bases; sum-of-contigs Quality
coverage: 6.41x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15843: contig of 15843 bp in length
* 15844 15943: gap of 100 bp
* 15944 36523: contig of 20580 bp in length
* 36524 36623: gap of 100 bp
* 36624 67681: contig of 31058 bp in length
* 67682 67781: gap of 100 bp
* 67782 119925: contig of 52144 bp in length
* 119926 120025: gap of 100 bp
* 120026 162689: contig of 42664 bp in length

```

```

* 162690 162789: gap of 100 bp
* 162790 174915: contig of 12126 bp in length
* 174916 175015: gap of 100 bp
* 175016 193301: contig of 18286 bp in length
* 193302 193401: gap of 100 bp
* 193402 204881: contig of 11480 bp in length.
FEATURES
Location/Qualifiers
source
1..204881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-477J21"
/clone_lib="RPC1-11.2"
1..15843
/note="assembly_fragment:00222"
vector_side:left"
vector_end:17
15944..36523
/note="assembly_fragment:02927"
36624..67681
/note="assembly_fragment:00442"
fragment_chain:1"
67782..119925
/note="assembly_fragment:02072"
fragment_chain:1"
120026..162689
/note="assembly_fragment:03061"
fragment_chain:1"
162790..174915
/note="assembly_fragment:00959"
fragment_chain:1"
175016..193301
/note="assembly_fragment:01787"
fragment_chain:1"
193402..204881
/note="assembly_fragment:01691"
fragment_chain:1"
vector_end:SP6
vector_side:right"
BASE COUNT 57526 a 45247 c 44379 g 57028 t 701 others
ORIGIN

Query Match 86.1%; Score 19.8; DB 81; Length 204881;
Best Local Similarity 91.3%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ctgccagccggaagcctggga 23
|||||
Db 39729 CTGCCAGCCCAAGGCTGGAA 39751
|||||

RESULT 6
AC006558/c 207392 bp DNA HTG 10-FEB-1999
LOCUS Homo sapiens chromosome 9 clone hRPK.477_J_21 map 9, *** SEQUENCING
DEFINITION IN PROGRESS ***, 7 unordered pieces.
ACCESSION AC006558
VERSION AC006558.1 GI:4262259
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 207392)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 9, clone hRPK.477_J_21
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207392)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,B., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,

```

Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McCurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Nilloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, J., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
Direct Submission
Submitted (10-FEB-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

TITLE
JOURNAL
COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3313: contig of 3313 bp in length
* gap of unknown length
* 3314 9632: contig of 6319 bp in length
* gap of unknown length
* 9633 24168: contig of 14536 bp in length
* gap of unknown length
* 24169 38366: contig of 14198 bp in length
* gap of unknown length
* 38367 56172: contig of 17806 bp in length
* gap of unknown length
* 56173 87103: contig of 30931 bp in length
* gap of unknown length
* 87104 207392: contig of 120289 bp in length.

FEATURES
source

Location/Qualifiers
1. .207392
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hrpk-477-J-21"
/clone.lib="RPC1-11 human BAC library"
/map="g"
/chromosome="9"

BASE COUNT 57991 a 44607 c 45345 g 58166 t 1283 others
ORIGIN

Query Match 86.1%; Score 19.8; DB 60; Length 207392;
Best Local Similarity 91.3%; Pred. No. 68;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ctgccagggcgaaggcctggga 23
||||||| |||||
Db 70048 CTGCCAGGCCAAGCCTGGAA 70026

RESULT 7
AE004779 15982 bp DNA BCT 30-AUG-2000
LOCUS
DEFINITION Pseudomonas aeruginosa PA01, section 340 of 529 of the complete genome.
ACCESSION AE004779 AE004091
VERSION AE004779.1 GI:9949735
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 15982)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,

Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
2 (bases 1 to 15982)
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
20437337
Location/Qualifiers
1. .15982
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
163. .1002
/gene="glpP"
/note="PA3581"
163. .1002
/gene="glpP"
/codon_start=1
/transl_table=11
/product="glycerol uptake facilitator protein"
/protein_id="AAG06969.1"
/db_xref="GI:9949736"
/translation="MTTAAPTPSLFGQCLAEFLGTALLIFFGTCVAAKLVAGASPLG
WEISLIMGVGSMAIYLSAGVSGAHLNPAVSLALWLFAGPEGKLPFYITQAVAGAF
AAALVYLYSLFIEFEQANIVRGSDSLASVSTYPHPALSYQAFVLEVVITA
ILMAVIMALTDGNGDLRGPLAPLILGLLIAGVIGSAMPLTGFMAMPADFGPKLMTY
LAGWGTAFTGGREIPYFLVPIFAPILGACLGAGGYRVLIARILPNSAAAPAEPEKV
RAS"
1042. .2559
/gene="glpK"
/note="PA3582"
1042. .2559
/gene="glpK"
/codon_start=1
/transl_table=11
/product="glycerol kinase"
/protein_id="AAG06970.1"
/db_xref="GI:9949737"
/translation="MTDKHNKYYVVALDQGTSSRAIVFDNDANVVSQAREFAOFYP
QAGVHDPEIWAQSTLVEALQASIEHQVAAIGITNORETTVWDRHSGRPIH
NAIVQCRSAAICAOIKRDGLDYIRETTGLVTDYFSGTKLKLILNVEGAREAR
NGDLFGTIDTLIWLITEKVHTDYTNASRTMLNHSRWDARMLEVLDPKSM
PEVNSSEYVGNARIGVGGELPIAGIAGDQQAALFGMCVEPQAKNTYGTGCFLL
MHTGDKAVSTHGLITIIACGPGVEGYALEGAVFNGSTVQVLRDELKVINDSFSE
YFAIKVDKSNVYLPFAFTGICAPYWDYARGAVFLTRGVKADHLIRATLESIAVQT
RDVDAQDAGERLARLVGDGAVANFQMFOADILGTRVVRPMRETTALGAAYL
AGLACGFWSLDELKSKAVIERVFECEDEPREKLYAGWKKAVERTGDDGEL"
2765. .3520
/gene="glpR"
/note="PA3583"
2765. .3520
/gene="glpR"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate regulon repressor"
/protein_id="AAG06971.1"
/db_xref="GI:9949738"
/translation="MNLPRQOQSIIEVRERGYLSIEEMAQMFVTPOTIRRDINOLA
EGGLRHYHGGAAVSSISNTAYTMRADMRDEKORIAEAVASLVPPNASILFINGT
TEALRALLNHRNLKIITNNLHVAATLSAKEDFEVLVAGCTVRSDGGIVGQAVDFIQ

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE
JOURNAL

FEATURES
source

gene

CDS

gene

CDS

gene

CDS

QFRVDFALVGIISGIDEDGSLDDFDYOEVRVSOAIIIDNARQVFLAADSSKFGNNAVRL
GPTSLNVRVISISPPPAALARLMAQHKVHLIV"
3798..5336
/gene="g1pD"
/note="PA3584"
3798..5336
/gene="g1pD"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate dehydrogenase"
/protein_id="AAG06972.1"
/db_xref="GI:9949739"
/translation="MSQHTPSAPLAEVVDVAVVGGINGVGTIAADAAGRLSVPLCE
QHDLAQHTSASSKLIHGGLRYLHEFRVLRALAEVLLAKAPHIVKPLRVFLPH
RPHLRPAMWIRAGFLYDHLGKREKLPASRGFLTGCTSPKAEIRRGFFYSCAVDDA
RLVNLAIASAREGAHVHTRTCVSARRSKGLWHLHRSKGLSLYIRARALVNAAGP
WVARFIDDLKQKSPYIRLIGOSHLIIVPKLYEGEHAYILONEDRIVEAIPYLDORFT
MIGTTDREYDQAKVAISEETAYILQVYNHFKQOLAADILHSFAGVPLCDDES
DEPSAITRDYTLSLAGNPEPLLSVFGGLTKITIRKLASALTQLQFPFANLGPWTA
KAPLPGEQMSQVEALTEOLANRYAMLRELARWARTYGTVRWRLDGVNGEADLGE
HLGGLYAREVDYLCKHEWAQDAEDILWRRSKLGFLSTSPQQVRLGQYLQSEPHRPR
VHAA"
5469..5798
/gene="g1pM"
/note="PA3585"
5469..5798
/gene="g1pM"
/codon_start=1
/transl_table=11
/product="membrane protein G1pM"
/protein_id="AAG06973.1"
/db_xref="GI:9949740"
/translation="MFKALIGAAVWLLAVLSKTRNYYIAGLVPPLPFTFALIAHYLVG
KGRSLDDKTTIVFGMWSIIPFYVYLAALYLLVERFRLETSALALAWLVAASVLVG
LWVRLHA"
complement(5815..6801)
/gene="PA3586"
complement(5815..6801)
/gene="PA3586"
/codon_start=1
/transl_table=11
/product="probable hydrolase"
/protein_id="AAG06974.1"
/db_xref="GI:9949741"
/translation="MRAEFLAALICSLSSFAAACDARVPTETVELGDVRLAYQSIG
RDSFALLVGLGQLHWEDEVVYVYALCEGFGFVIRYDNEVGLSANNVVPSSRLT
YEVVRYRLGLPWSADYTLTDWAGDALHLLDALDIPQAVILGASGGMLAQHIAADPAQ
RLSLTLVMTSGAGELPAPSESLRLLARREASREQAVEQADLLAAGSPVPRND
ROOLLQARSYDRAFNPEGVORQLLATLAPSRVPLNRLVOPTLVIHGTADPLLPV
MHGVHAAHISYRSELKLTPLGLAHRFQEAKEPFLAAVVPYILKAHRAQSVVQL"
7100..8020
/gene="metR"
/note="PA3587"
7100..8020
/gene="metR"
/codon_start=1
/transl_table=11
/product="transcriptional regulator MetR"
/protein_id="AAG06975.1"
/db_xref="GI:9949742"
/translation="MLELRHLKTLHALREADSLVAAERLHLTQSALSHQPKLEERL
GMOLFVRKTPVRFVTAGLRLQLADSLPLQRTAERDLARIAGGTAGRLHMAIECHP
CFOWLMPTIDQFRAWPEVELDLAGFSFAPALARGDLVLVTFADVEVLPFTIYVP
LFTYVALLAVNOHALAGRPYVPEPDERETLITYPVERDRDLDTFTFLDPADVPAPQ
VRIAEITYMMQLVASGRVCCCLPNWALHEYSRGVYVITARRLGERGLYCTLFAAIRTD
MEDAPFMRDPLTAKTISPATLEGVSAAG"
complement(8026..9276)
/gene="PA3588"
complement(8026..9276)
/gene="PA3588"
/codon_start=1
/transl_table=11
/product="probable porin"

/protein_id="AAG06976.1"
/db_xref="GI:9949743"
/translation="MNNKKAPELPLILAGDSVPALAEGLFEDSRASLALRNFYMR
DFRAGRAKSEWAQGFLEFDRSGYTEGLTGLVGLDLGLKLVRLDSGASRGTLPLP
LRDDGSAAGDYARLDSRSELKLVGLPKLPTIQPNVGRFLPPOVFOGALLTS
GELSCSLNGLRUTEVSORNEAGTSDIALFNRRNRFAAGADRFLDGLADYRIAPDW
TGSYHGELEQVYAHLEGLKIGIADSLIESDLRLALSRTDGGAGRGIRDNKFSFG
SLTYRLNGOAFGIGYQRMGSDHGFYPLEGTDPLYLVNFQYNDFAEGSSWQRLYDC
DFAPLGVPLSLMTRFESGHCAKPKGAGDSREWERDSDLRVVLQGGALKGLGWRNA
TYRSFASRDIDENKLYLYELPLF"
complement(9329..10534)
/gene="PA3589"
complement(9329..10534)
/gene="PA3589"
/codon_start=1
/transl_table=11
/product="probable acyl-CoA thiolase"
/protein_id="AAG06977.1"
/db_xref="GI:9949744"
/translation="MLDAYIYAGLTPFRHAGALSTVRPDDLAGLLRLARLAETSGFA
VDDLEDTLGTNOAGDSRNLARNALLAAGFARLPQGTVNRILCASLSAVIDAARA
TSCGRLYLAGGAEFMSRPAEVMGKAFSAFRTLEVFDSTIGARFANPRIVRGND
SMPTGVNVARAFGIAREDAEDFAASSQARYQAALEGEFFLGEILPVEVRAGRGETR
LIVERDEHPRPQADLAALRLALFALFAGGVVTFAGNAGSINGDRAVLLGDRAIGREGIR
PLARILASASGVPEPLMGIGPQQAILRALORAGIDDEVLGLIELINEAFAPQVLACLK
LIGLDYEDPRVYNPHGGATLGHPLGASGARLVLTAAARGLQRIERRIYVAVSLCVGLGQC
VAMVIERCR"
complement(10557..12089)
/gene="PA3590"
complement(10557..12089)
/gene="PA3590"
/codon_start=1
/transl_table=11
/product="probable hydroxyacyl-CoA dehydrogenase"
/protein_id="AAG06978.1"
/db_xref="GI:9949745"
/translation="MSEPSIVLIGIVGTGAMGGIAQLAACAGLSVLLYDSRQGVALQ
AREQIATVLAQVGRLEAEAEVAMGNLYVEDLVLVGLGQVLEAVELEAKQA
LRFQLEEVVYDEAILASNTSSVTAISACRDPRIAGLHFFNPVPLMRIVEVIEGL
ATRTGIAERLCAVATFQHVAVRTDPSGFIVNHAGRAGFTGALRILGECYAPVAID
EVLREGAGFMGPFELDLVGLDVLSPVMESTYRQYEEPRYRPHPLRLQMLAAGRLG
RKSGOGFYRYDAGQVPAAPVAPGAALPPVVLGVDDDEHRAPLMLLQRLGAEVES

Query Match 82.6%; Score 19; DB 1; Length 15982;
Best Local Similarity 100.0%; Pred. No. 4.2e+02; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0;
Qy 3 gccacagccgaagcctgg 21
|||||
Db 8404 GCCACGCCGAGGCCCTGG 8422
|||||
RESULT 8
AC073464 169194 bp DNA HTG 15-JUL-2000
LOCUS Homo sapiens chromosome 2 clone RP11-134N21, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
AC073464
AC073464.2 GI:8954226
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
2 (bases 1 to 169194)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 169194)
Waterston, R.H.
Direct Submission
Submitted (18-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8571786.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: HNH0134N21
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 160139 bases at least Q40
Consensus quality: 162624 bases at least Q30
Consensus quality: 163974 bases at least Q20
Insert size: 166000; agarose-fp
Quality coverage: 4.34 in Q20 bases; agarose-fp
Quality coverage: 4.36 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1364: contig of 1364 bp in length
* 1365 1464: gap of unknown length
* 1465 3740: contig of 2276 bp in length
* 3741 3840: gap of unknown length
* 3841 7587: contig of 3747 bp in length
* 7588 7687: gap of unknown length
* 7688 11136: contig of 3449 bp in length
* 11137 11236: gap of unknown length
* 11237 14308: contig of 3072 bp in length
* 14309 14408: gap of unknown length
* 14409 21712: contig of 7304 bp in length
* 21713 21812: gap of unknown length
* 21813 27149: contig of 5337 bp in length
* 27150 27249: gap of unknown length
* 27250 35937: contig of 8688 bp in length
* 35938 36037: gap of unknown length
* 36038 47252: contig of 11215 bp in length
* 47253 47352: gap of unknown length
* 47353 58329: contig of 10977 bp in length
* 58330 58429: gap of unknown length
* 58430 72793: contig of 14364 bp in length
* 72794 72893: gap of unknown length
* 72894 85444: contig of 12551 bp in length
* 85445 103459: gap of unknown length
* 103460 103559: gap of 17915 bp in length
* 103560 124591: contig of 21032 bp in length
* 124592 147042: gap of unknown length
* 147043 147142: contig of 22351 bp in length
* 147043 169194: gap of unknown length
* 147143 169194: contig of 22052 bp in length.

FEATURES
source
1. 169194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-134N21"
1. 1364
/note="assembly_name:Contig4"
1465. 3740
/note="assembly_name:Contig5"
3841. 7587

misc_feature
misc_feature
misc_feature

misc_feature /note="assembly_name:Contig6"
7688. 11136
misc_feature /note="assembly_name:Contig7"
11237. 14308
misc_feature /note="assembly_name:Contig8"
14409. 21712
misc_feature /note="assembly_name:Contig9"
21813. 27149
misc_feature /note="assembly_name:Contig10"
27250. 35937
misc_feature /note="assembly_name:Contig11"
36038. 47252
misc_feature /note="assembly_name:Contig12"
47353. 58329
misc_feature /note="assembly_name:Contig13"
58430. 72793
misc_feature /note="assembly_name:Contig14"
72894. 85444
misc_feature /note="assembly_name:Contig15"
clone_end:17
vector_side:right
85545. 103459
misc_feature /note="assembly_name:Contig16"
103560. 124591
misc_feature /note="assembly_name:Contig17"
124692. 147042
misc_feature /note="assembly_name:Contig18"
147143. 169194
misc_feature /note="assembly_name:Contig19"
BASE COUNT 51420 a 32271 c 33139 g 50860 t 1504 others
ORIGIN

Query Match 82.6%; Score 19; DB 74; Length 169194;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gccacggccgaaggcctgg 21
|||||
Db 43409 GCCCAGGCGGAGGCGCTGG 43427

RESULT 9
AL137070/c
LOCUS Human DNA sequence from clone RP11-251017 on chromosome 9, complete
DEFINITION AL137070.9 GI:9367917
ACCESSION AL137070
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170237)
AUTHORS Skuce,C.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 22, 2000 this sequence version replaced gi:8977609.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

Thu Oct 25 13:09:01 2001

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Embl, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-251017 is from the library RPI-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-251017.

```

FEATURES             source
    source
        1..170237
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="9"
        /clone="RP11-251017"
        /clone_lib="RPI-11.1"
        /note="Random repeat. Single clone region. Assembly confirmed by restriction enzyme digest data."
    misc_feature
        98259..98301
        /note="Tandem repeat. Short join in tandem repeat confirmed by ECORI and HINDIII digest data."
    misc_feature
        98302..98403
        /note="Tandem repeat. Single clone region. Assembly confirmed by restriction enzyme digest data."
BASE COUNT          54028 a 32613 c 32057 g 51539 t
ORIGIN
Query Match          82.6%; Score 19; DB 89; Length 170237;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gccacggccgaaggcctgg 21
|||||
DB 26857 GCCACGGCCGAAGGCTGG 26839

RESULT 10
AF151965 5162 bp DNA BCT 01-JUN-2000
LOCUS Pseudomonas putida strain KT2440 proline dehydrogenase (putA) gene,
DEFINITION complete cds.
ACCESSION AF151965
VERSION AF151965.1 GI:8132051
KEYWORDS Pseudomonas putida.
SOURCE Pseudomonas putida
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 5162)
Vilchez, S., Molina, L., Ramos, C. and Ramos, J.L.
Proline catabolism by Pseudomonas putida: cloning, characterization
and expression of the put genes
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 5162)
Vilchez, S., Molina, L., Ramos, C. and Ramos, J.L.
Direct Submission
TITLE Submitted (16-MAY-1999) Biochemistry, CSIC, Profesor Albareda No.1,
JOURNAL Granada E-18008, Spain
FEATURES
source
1..5162
/organism="Pseudomonas putida"
/db_xref="taxon:303"
/strain="KT2440"
359..4306
/gene="putA"
359..4306
CDS

```

```

/gene="putA"
/note="PutA; proline oxidase; pyrroline 5'carboxylate
dehydrogenase"
/codon_start=1
/transl_table=11
/product="proline dehydrogenase"
/protein_id="AAF73193.1"
/db_xref="GI:8132052"
/translation="MATTLGVKLDPTRELKAAOSIDRPHLLIKQAIQVNLKEL
EGGATLTENGLHASPADDAGEVQADSHQCFLEPAESILPOSVLSAITAAYRRPQ
EYVPMLEQLRASPADATNKLAAETAKLRNOKSVGGAGIVQGLLOQFSLSSQER
RGVCLAEALLRIPDRKTDALIRDKISTGNQPHLGNPSFLVNAATWGLLTGKLVG
THNETGLTSLTRIICKSGEPMIRKVDMAIMLGEQFVTGETIAEALANASPEAKG
FRYSQVDMLEAALTEHDAQKYLASVEQAIHSIGKASHGRGIYEGPGISIKLSALHPRY
SRAQYRMEELYPRLLSILIAKOYDIGLINDAEADRLSLDLRLERLCFEPFLAG
WNGIGVQAYQKRCPYLLINFEFLAKRTPHRLIRLLKGYMDELKRAQVEGLECY
PVYTRKYDVSIVACAKKLLAPYPAFYQFATHNAHTLSAIYHAGONYPGQYEFQ
CLHGMGEPLYEQVVGKIADKLNRCRVYAPVGTHTLALLVRLLENGANTSFVNR
IADHSISIQELVADRWPAIAWVPRKSGISGLPHRIPRLFROLIYTERAKLAGIEMANE
HRLGLISCAMVATAHKQWEAAPLLACAAASAAPVLPADPNRVNVGHVQATVAKFD
NATHCALNPATWQATPPAARAAIILERTADLMEAEIHPIMGLLTREAGKTFPNAIEL
REAVDRLVYAVQALNDFSNDARPLGPVVCISPNWFLPALTGTQVAAALAGNPVLA
KBAOTPLIAQAQVRLLEAGIPGVLQLLPGRGTVGAGLVGDERVKNVFTGSTEV
ARLLQNWAGLDNRVIELKGMAMAEISGCDRLAVDIGVIDAEAKAGIEKHIOGM
REKREYVQVATADAAEIKRGTFVMTLIEDSFDELKREIFGPVLHVRYNRNLDQ
LIEQINNSGYLTLGVHTRIDETIAKVETATPATCRHNIIVGAVGVQPPGGGLSG
TGPAGGPLYLYRLLTTPADAIGRHFOQDGEPTDRTLHEQLVKPLHGLKAWAENN
QADLAALCSQFASQSGIARLLPCTGCRNSYITLPREHVLCLADNETDLLAFAA
VLAVSSAVNVQDGECKAIRLPRELQAKVKLVADWNKDEVAFDAVTHHGDSQLRG
VCOQVAKRAGAIQVGHGSSGDRHQLERLVIERSVNTAAAGGNASLMTIG"
BASE COUNT          1051 a 1702 c 1502 g 907 t
ORIGIN
Query Match          81.7%; Score 18.8; DB 2; Length 5162;
Best Local Similarity 90.9%; Pred. No. 8.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tqccacggccgaaggcctggga 23
|||||
DB 1681 TCCACAGCTCGAAGGCTGGAA 1702

RESULT 11
AF153207/c 8660 bp DNA BCT 16-JAN-2000
LOCUS Pseudomonas putida proline dehydrogenase (putA) and proline
DEFINITION permease (putP) genes, complete cds.
ACCESSION AF153207
VERSION AF153207.1 GI:5693828
KEYWORDS Pseudomonas putida.
SOURCE Pseudomonas putida
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 8660)
Vilchez, S., Molina, L., Ramos, C. and Ramos, J.L.
Proline catabolism by pseudomonas putida: cloning,
characterization, and expression of the put genes, in the presence
of root exudates
J. Bacteriol. 182 (1), 91-99 (2000)
JOURNAL
MEDLINE
20082847
10613867
REFERENCE
2 (bases 1 to 8660)
Vilchez, S., Molina, L., Ramos, C. and Ramos, J.L.
Direct Submission
TITLE Submitted (20-MAY-1999) Biochemistry, CSIC, Profesor Albareda No.1,
JOURNAL Granada E-18008, Spain
FEATURES
source
1..8660
/organism="Pseudomonas putida"
/strain="KT2440"

```

```

/db_xref="taxon:303"
complement(857..4804)
/gene="putA"
CDS
complement(857..4804)
/gene="putA"
/codon_start=1
/transl_table=11
/product="proline dehydrogenase"
/db_xref="GI:6693829"
/translation="MATTTILGVKLDPTRELRKAAQSIDRTPHLWIKQAIIFYLEKIL
EGATITELNGHASNPADAGEVQADHSQCLEFAESILQSVLRSIAITAYRRPEO
EVPMLLEQARLUSADATANKLASIAEKLKQSVGGVGRAGIVGLQLEFSLSSQER
RGVCLAEALLRTPDKGTRDALRDKISTGNMOPHLGNSPSLFVNAATWGLLITGLKLS
THNETGITSLLRIIGKSGEPMIRKGVDMAMPELQEVGTETIAEALANSPREAKG
FRYSYDMGCEALYTHDAOKYASVEOATHSGKASHGSGYEGEGISIKUSALHPRY
SRAOYERMEELYPRLLSITLAKQYDITGLNDIAEADRLSDLDLERICFEPSLAG
WNGIGVIAQKRPYLNIFFDLAKRTPHRLIIRLLKGYWDSITKRAQVELEGY
PVTKRIVDSIVACAKKLAVPEALYPOFATHNAHTLSAIYHAGNYPGQYEFQ
CLHGMEPLYEQVQKADGKI.NRRCRVYAVGTHETLLAVLRLLENGANTSFNVR
IADHSIQELVADRWPAISIAWPKRSGISGLPHRPIPLRDLGYTERAKLAGIFMARE
HRLGLSCAMVATKOWEAPALLACAARESAAAPVLNADHRNVVGHVQVATVAKED
NAIHCALNPATWQATPPAERAAIILERTADI.MEAEIHPMLGILLIREAKTFFPNAIEL
REAVDFLRVYVAQALNDFSNDRHRLGPVCLSPWNFPPLAIFTGVOAAALAGNPVLA
KPAEQTPLIAQAVRLLLEAGIEGVQLLIPGRGETVAGLVGDERVKGYMFTGSTEV
ARLLQNNVAGRLDNGRPTPLIAETGSGONAMIVDSALLTEQWIDVYSSAFDSAGQRC
SALRVLICQEDSADRVIELMGAMAESRLGCPDLAVDIGPVIDAEAKAGIEKHIOGM
REKGRPVQVALADAAEIKRGTFVMTPLIELDSDELKRETFGLVPLHVYNNRLDO
LIEOINNSGYTLTGVTTRIDETAKVVEATPATCRRNIVGAVGVQPPGEGELSG
TGPKAGGPLYLVRLSTRPADAIGRHFQQQDGEPTDRLHEQLKPLHLGKAWAENN
OLADUALCSQFASQSGIARLIPGPTGERNYSYIIPREHVLCLADNETDLLAQFAA
VLAVGSSVDVDEGPKALRPRELOAKYKLVADWNKDDEVADFVIRHGDSDQLRG
VCCQVAKRAGATVGVHGLSSGDHQIALERLIERAVSVNTAAAGGNASLMTIG"
5160..6639
/gene="putP"
CDS
5160..6638
/gene="putP"
/codon_start=1
/transl_table=11
/product="proline permease"
/protein_id="AAF25001.1"
/db_xref="GI:6693830"
/translation="MGNPLTITFVIYITAMVILGFAYRATNMLSDYLIGRSLGSVV
TALASGASDMGMLLMLGPGAIYFAGLSEAWIAGITVGAYLNLVAGRLRVQTEHN
GDALTPPYFSSREDNSGLLRIISAIVILVFTTCASGIVAGARLFESTFGMSYET
ALWAGRAIATYTVGGFLAVSMTDITVQASLMIFALITPVIILSTGGDFDTFAAIE
AVNRNFDLKGATFIGIISLMGWLGFYFRPHILARFMAADSVNIAKARRISMTWM
ILCLAGTCVAGCFIAYFSAHPPELAGPVSENHKRVFTIELAKILFNPNVAGVLLSALLA
AMVITLSCQLLCVCSALTENFYKAFIRKNASOVELVWGLVLAVALIAIAMAANPE
NRVLGLVAYAWAGFGAAGFPVLLISLVKMGTRNGALAGLVVGVGALTVILWKNFDTLGL
YEIIPGFLFASIAIVLSVLSKGSQTHVRFKFEADAAYHADK"
BASE COUNT 1563 a 2590 c 2748 g 1759 t
ORIGIN

Query Match
Best Local Similarity 81.7%: Score 18.8; DB 2; Length 8660;
Matches 20: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 tggccaggccgaagcctggga 23
||||||| |||||||||
Db 3482 TGCCAGGTCGAAGCCTGGAA 3461

RESULT 12
LOCUS SCF1 14750 bp DNA BCT 09-SEP-1999
DEFINITION Streptomyces coelicolor cosmid F1.
ACCESSION AL117322
VERSION AL117322.1 GI:5869937
KEYWORDS acetyltransferase; alcohol dehydrogenase; binding-protein-dependent
transport; hydrolase; hydroxylase; LysR-family; MarR family;
oxidoreductase; transcriptional regulator.

```

Streptomyces coelicolor A3(2).
Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycetaceae; Streptomyces.
1 (bases 1 to 14750)
Redenbach, M., Kieser, H.M., Denapalite, D., Eichner, A., Cullum, J.,
Kinashi, H. and Hopwood, D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
2 (bases 1 to 14750)
Seeger, K.J. and Harris, D.
Unpublished
3 (bases 1 to 14750)
Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (09-SEP-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/projects/S.coelicolor/)
CDS are numbered using the following system eg SC7B7.01c, SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.gov/jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid F1 Lies between and overlaps with cosmids F20 and F85 on the
AseI-F genomic restriction fragment.
Location/Qualifiers
1..14750
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid F1"
complement(1..187)
/note="Nominal overlap with Streptomyces coelicolor cosmid
F20."
c1..505
/gene="SCF1.01"
/note="SCF1.01, partial CDS, probable alcohol
dehydrogenase, len: >167aa; similar the C-terminal region
of many eg. SW:ADH3_SULSO NAD-dependent alcohol
dehydrogenase from Sulfolobus solfataricus (347 aa) fasta
scores; opt: 274, z-score: 304.7, E(): 1.3e-09, (31.8%
identity in 170 aa overlap). Also similar to C-terminal
region of TR:BAA82700 (EMBL:AB017438) from Streptomyces


```

coelicolor (346 aa) fasta scores: opt: 372, z-score:
409.3, E(): 2e-15, (40.7% identity in 167 aa overlap).
/codon_start=2
/transl_table=11
/label=SCF1.01
/product="putative alcohol dehydrogenase"
/protein_id="CAB55521.1"
/db_xref="GI:5869938"
/translation="IGAGLGHIGIOLGALTAETIVVDNRNPDAVDLAVSGADHGV
LADGHVDRVRELTTGGHAEAVDFVGGEGTGDGICMLRRAGDYHVVGGENIDVPT
LDVTSAPFINFIGNVGSYNDLCELMVLAARGVRRLTKAKYSLDRFDQALDDGGRIR
GRALVP"
1. 505
/gene="SCF1.01"
558. .887
/gene="SCF1.02"
558. .887
/gene="SCF1.02"
/note="SCF1.02, unknown, len: 109aa; similar to TR:006569
(EMBL:295385) hypothetical protein from Mycobacterium
tuberculosis (107 aa) fasta scores: opt: 433, z-score:
595.2, E(): 8.7e-26, (56.6% identity in 106 aa overlap)."
/codon_start=1
/transl_table=11
/label=SCF1.02
/product="hypothetical protein"
/protein_id="CAB55522.1"
/db_xref="GI:5869939"
/translation="MIFITAKFRVLPHEADQWPQVEDFTTRATRAEPGLWFQWSRL
DDPEYVLVEAFRDEAGAAHVGSAAFKAAQQTLPPLHATTPTRVNANVPQDDWSLLG
EMAYDAG"
complement(967. .1494)
/gene="SCF1.03c"
complement(967. .1494)
/gene="SCF1.03c"
/note="SCF1.03c, probable acetyltransferase, len: 175aa;
similar to many eg. SW:MAA_ECOLI maltose
o-acetyltransferase from Escherichia coli (182 aa) fasta
scores: opt: 365, z-score: 432.1, E(): 1.1e-16, (43.6%
identity in 140 aa overlap). Contains Pfam match to entry
PF00132 hexapep, Bacterial transferase hexapeptide (four
repeats) and Prosite match to PS00101 Hexapeptide-repeat
containing-transferases signature."
/codon_start=1
/transl_table=11
/label=SCF1.03c
/product="putative acetyltransferase"
/protein_id="CAB55523.1"
/db_xref="GI:5869940"
/translation="WVHTPEFARHAERIVEVTDATSLNVLFPFSDSAGRSELLSVVF
GGPLPESVTIYPPFTTGLNTFCGIVFVNOGCTFMDKGGIRIGNRVMIAPKASLVT
GGHPLPLARRAHLSPAPIVEDDVIWIGTAAVITQGVIGAGVVAAGAVVTRDVPAG
TVVAGVPARVHKQIG"
complement(1024. .1128)
/gene="SCF1.03c"
/note="Pfam match to entry PF00132 hexapep, Bacterial
transferase hexapeptide (four repeats), score 49.50,
E-value 3.2e-12"
complement(1024. .1110)
/gene="SCF1.03c"
/note="PS00101 Hexapeptide-repeat containing-transferases
signature."
1608. .2543
/gene="SCF1.04"
1608. .2543
/gene="SCF1.04"
/note="SCF1.04, probable LysR-family transcriptional
regulator, len: 311aa; similar to many transcriptional
regulators especially in the N-terminal region eg.
SW:OXYR_MYCAV regulator of hydrogen peroxide responsive
genes from Mycobacterium avium (311 aa) fasta scores: opt:
292, z-score: 329.4, E(): 5.5e-11, (36.6% identity in 191
aa overlap). Contains Pfam match to entry PF00126 HTH_1,

```

```

Bacterial regulatory helix-turn-helix protein, lysR family
and prosite match to PS00044 Bacterial regulatory
proteins, lysR family signature."
/codon_start=1
/transl_table=11
/label=SCF1.04
/product="putative LysR-family transcriptional regulator"
/protein_id="CAB55524.1"
/db_xref="GI:5869941"
/translation="MDVEPALRTFVVAGVGQFAAADDELGISQOAVSKRIAALRRHVA
VTLVRSRGRSLSDGGVFLPHAKKVLTAVEQAVNPGSRPLRVDVLRNRRISPAQ
AVYRYSRPEPTDLDVADTKENAAQAQVLEGVDASFALPADRVPTCITAEIRLL
DEPLELLVPGPHLADAPRISEVDLAGHRIWIPGIRPCTEWAATFQALSEAFGLSIDA
LCPNFGDEALMDALSDASLATLVGRDRLVPRTHDLRRLPLHDPFPFPHLLLFRT
GDRHPVLTALRDHLRTTAPRTPHDAWTPDWTVTVH"
1614. .2033
/gene="SCF1.04"
/note="Pfam match to entry PF00126 HTH_1, Bacterial
regulatory helix-turn-helix protein, lysR family, score
83.50, E-value 4.3e-21"
1656. .1748
/gene="SCF1.04"
/note="PS00044 Bacterial regulatory proteins, lysR family
signature."
complement(2567. .3682)
/gene="SCF1.05c"
complement(2567. .3682)
/gene="SCF1.05c"
/note="SCF1.05c, possible oxidoreductase, len: 371aa;
weakly similar to many eg. TR:069945 (EMBL:AL023862)
putative oxidoreductase from Streptomyces coelicolor (430
aa) fasta scores: opt: 224, z-score: 247.7, E(): 2e-06, to
(26.0% identity in 392 aa overlap). Contains Pfam match to
entry PF01408 GFO_IDH_MocA, Oxidoreductase family."
/codon_start=1
/transl_table=11
/label=SCF1.05c
/product="putative oxidoreductase"
/protein_id="CAB55525.1"
/db_xref="GI:5869942"
/translation="MNTPGSALADLARSGRPIGVGIGLGAHGSWAERSHLPALRAVG
GFELRALSTSSKQSAERKHCVRRAFGTAELACDEVDLVVAVKVPHREILVRT
VLNAGKSLCEPPLGNLAEPEDMARLARAAVPTVVGLOARSHDPAFAYTRHLVADGY
VGEVLSTTVVGGGANGAVGPADHVLDAANGATLLTVFGHALDGVASVLGPEDEL
RIOAASRRTSAVDITETGRVPVMTAQQVWASGRLPGAVATPHYGSGSRGTFNFEWI
NGTGDVLTPVPGHLPQLSPLTLEGGRTSLALAPITVPEAYARVARLEPRADAPAYA
VAHAYQRFLLDLREGTAHVDPDFAHGVVHRHSIEQYLP"
Query Match 81.7% Score 18.8; DB 3; Length 14750;
Best Local Similarity 90.9%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 tgcacagccgcaaggcctggga 23
||| ||||| ||||| ||||| |||
Db 2253 TGCTCAGGCCGCAAGGCTTCGGA 2232

RESULT 13
AP001078 176253 bp DNA HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-861B19 map 18p11.3, WORKING
DEFINITION DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AP001078
VERSION AP001078.2 GI:8117771
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 176253)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,I., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 176,253 genomic DNA of 18p11.3

```


JOURNAL
REFERENCE
AUTHORS

Published Only in Database (2000) In press
2 (bases 1 to 176253)

Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Tokoki, Y., Watanabe, H., and Sakaki, Y.
Direct Submission
Submitted (17-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail: hattori@gsc.riken.go.jp)
URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-42-778-9923,
Fax: 81-42-778-9924

COMMENT

On May 31, 2000 this sequence version replaced gi:6997425.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-861B19

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 155758 bases at least Q40

Consensus quality: 165815 bases at least Q30

Consensus quality: 170532 bases at least Q20

Insert size: 173753; sum-of-contigs

Quality coverage: 4.13x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 25481 contig of 25481 bp in length
25582 44141 contig of 18560 bp in length
44242 59207 contig of 14966 bp in length
59308 73163 contig of 13856 bp in length
73264 83431 contig of 10168 bp in length
83532 90995 contig of 7464 bp in length
91096 99746 contig of 8651 bp in length
99847 105997 contig of 6151 bp in length
106098 112950 contig of 5024 bp in length
113051 118074 contig of 5838 bp in length
118175 124012 contig of 5257 bp in length
124113 129369 contig of 5641 bp in length
129470 135110 contig of 5936 bp in length
135211 141146 contig of 5364 bp in length
141247 146610 contig of 4112 bp in length
146711 150822 contig of 4308 bp in length
150923 155230 contig of 3430 bp in length
155331 158760 contig of 2446 bp in length
158861 163143 contig of 4283 bp in length
163244 165114 contig of 1871 bp in length
165215 167768 contig of 2554 bp in length
167869 170314 contig of 100 bp in length
170415 171946 contig of 1532 bp in length
171947 172047 contig of 100 bp in length
172047 173708 contig of 1662 bp in length
173709 173809 contig of 100 bp in length
173809 174921 contig of 1113 bp in length
174922 175022 contig of 100 bp in length
175022 176253 contig of 1232 bp in length

```

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 25481: contig of 25481 bp in length

```

* 25482 25581: gap of 100 bp
* 25582 44141: gap of 18560 bp in length
* 44142 44241: gap of 100 bp
* 44242 59207: contig of 14966 bp in length
* 59208 59307: gap of 100 bp
* 59308 73163: contig of 13856 bp in length
* 73164 73263: gap of 100 bp
* 73264 83431: contig of 10168 bp in length
* 83432 83531: gap of 100 bp
* 83532 90995: contig of 7464 bp in length
* 90996 91095: gap of 100 bp
* 91096 99746: contig of 8651 bp in length
* 99747 99846: gap of 100 bp
* 99847 105997: contig of 6151 bp in length
* 105998 106097: gap of 100 bp
* 106098 112950: contig of 6853 bp in length
* 112951 113050: gap of 100 bp
* 113051 118074: contig of 5024 bp in length
* 118075 118174: gap of 100 bp
* 118175 124012: contig of 5838 bp in length
* 124013 124112: gap of 100 bp
* 124113 129369: contig of 5257 bp in length
* 129370 129469: gap of 100 bp
* 129470 135110: contig of 5641 bp in length
* 135111 135210: gap of 100 bp
* 135211 141146: contig of 5936 bp in length
* 141147 141246: gap of 100 bp
* 141247 146610: contig of 5364 bp in length
* 146611 146710: gap of 100 bp
* 146711 150822: contig of 4112 bp in length
* 150823 150922: gap of 100 bp
* 150923 155230: contig of 4308 bp in length
* 155231 155330: gap of 100 bp
* 155331 158760: contig of 3430 bp in length
* 158761 158860: gap of 100 bp
* 158861 163143: contig of 4283 bp in length
* 163144 163243: gap of 100 bp
* 163244 165114: contig of 1871 bp in length
* 165115 165214: gap of 100 bp
* 165215 167768: contig of 2554 bp in length
* 167769 167868: gap of 100 bp
* 167869 170314: contig of 2446 bp in length
* 170315 170414: gap of 100 bp
* 170415 171946: contig of 1532 bp in length
* 171947 172046: gap of 100 bp
* 172047 173708: contig of 1662 bp in length
* 173709 173808: gap of 100 bp
* 173809 174921: contig of 1113 bp in length
* 174922 175021: gap of 100 bp
* 175022 176253: contig of 1232 bp in length.

```

FEATURES
source

```

1..176253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-861B19"

```

misc_feature

```

1..25481
/notes="assembly_fragment"
25582..44141
/notes="assembly_fragment"
44242..59207
/notes="assembly_fragment"
59308..73163
/notes="assembly_fragment"
73264..83431
/notes="assembly_fragment"
83532..90995
/notes="assembly_fragment"
91096..99746
/notes="assembly_fragment"
99847..105997
/notes="assembly_fragment"

```

Thu Oct 25 13:09:01 2001

```

misc_feature 106098..112950
/notes="assembly_fragment clone_end:sp6 vector_side:left"
misc_feature 113051..118074
/notes="assembly_fragment"
misc_feature 118175..124012
/notes="assembly_fragment"
misc_feature 124113..129369
/notes="assembly_fragment"
misc_feature 129470..135110
/notes="assembly_fragment"
misc_feature 135211..141146
/notes="assembly_fragment"
misc_feature 141247..146610
/notes="assembly_fragment"
misc_feature 146711..150822
/notes="assembly_fragment"
misc_feature 150923..155230
/notes="assembly_fragment"
misc_feature 155331..158760
/notes="assembly_fragment"
misc_feature 158861..163143
/notes="assembly_fragment"
misc_feature 163244..165114
/notes="assembly_fragment"
misc_feature 165215..167768
/notes="assembly_fragment"
misc_feature 167869..170314
/notes="assembly_fragment"
misc_feature 170415..171946
/notes="assembly_fragment"
misc_feature 172047..173708
/notes="assembly_fragment"
misc_feature 173809..174921
/notes="assembly_fragment"
misc_feature 175022..176253
/notes="assembly_fragment"
BASE COUNT 48857 a 38855 c 39248 g 46787 t 2506 others
ORIGIN

```

```

Query Match . 81.7%; Score 18.8; DB 82; Length 176253;
Best Local Similarity 90.9%; Pred. No. 1.Re+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ctgccagccgcaagcctggg 22
Db 99569 CTGCCAAGGCCCAAGCCCTGGG 99590

```

```

RESULT 14
LOCUS AP001793 193487 bp DNA HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-661013 map 18p11.3, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP001793
VERSION AP001793.2 GI:8117465
KEYWORDS HTG; HTGS; PHASE1; HTGS-DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-661013.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 193487)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 193,487 genomic DNA of 18p11.3
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 193487)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

```

```

Japan (E-mail: hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7592907.
----- GENOME CENTER
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-661013
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 176068 bases at least Q40
Consensus quality: 185034 bases at least Q30
Consensus quality: 188618 bases at least Q20
Insert size: 190387; sum-of-contigs
Quality coverage: 4.33x in Q20 bases; sum-of-contigs
-----

```

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 15033 contig of 15033 bp in length
2 28850 contig of 13717 bp in length
3 15134 28951
4 43381 contig of 1431 bp in length
5 54311 contig of 10830 bp in length
6 55508 contig of 11097 bp in length
7 74474 contig of 8866 bp in length
8 86282 contig of 11708 bp in length
9 74575
10 94094 contig of 7712 bp in length
11 102440 contig of 8246 bp in length
12 109661 contig of 7121 bp in length
13 118186 contig of 8425 bp in length
14 118287 contig of 5023 bp in length
15 123309 contig of 5567 bp in length
16 123410 contig of 4908 bp in length
17 13984 contig of 4390 bp in length
18 134085
19 138575
20 149770
21 153637
22 158023 contig of 4387 bp in length
23 162445 contig of 4322 bp in length
24 158124
25 162546
26 165243 contig of 2698 bp in length
27 16344
28 168976
29 171973 contig of 2998 bp in length
30 175989 contig of 3916 bp in length
31 179781 contig of 3692 bp in length
32 182243 contig of 2362 bp in length
33 182344
34 184565 contig of 2222 bp in length
35 187147 contig of 2482 bp in length
36 189548 contig of 2301 bp in length
37 187248
38 189649
39 190661 contig of 1013 bp in length
40 190762
41 192068
42 193487 contig of 1420 bp in length
Sequence updated (26-May-2000).

```

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 15033: contig of 15033 bp in length

* 15034 15133: gap of 100 bp

* 15134 28850: contig of 13717 bp in length

*	28851	28950:	gap of	100 bp
*	28951	43381:	contig of 14431 bp in length	
*	43382	43481:	gap of	100 bp
*	43482	54311:	contig of 10830 bp in length	
*	54312	54411:	gap of	100 bp
*	54412	65508:	contig of 11097 bp in length	
*	65509	65608:	gap of	100 bp
*	74475	74474:	contig of 8866 bp in length	
*	74575	86282:	contig of 11708 bp in length	
*	86283	86382:	gap of	100 bp
*	86383	94094:	contig of 7712 bp in length	
*	94095	94194:	gap of	100 bp
*	94195	102440:	contig of 8246 bp in length	
*	102441	102540:	gap of	100 bp
*	102541	109661:	contig of 7121 bp in length	
*	109662	109761:	gap of	100 bp
*	109762	118186:	contig of 8425 bp in length	
*	118187	118286:	gap of	100 bp
*	118287	123309:	contig of 5023 bp in length	
*	123310	123409:	gap of	100 bp
*	123410	128976:	contig of 5567 bp in length	
*	128977	129076:	gap of	100 bp
*	129077	133984:	contig of 4908 bp in length	
*	133985	134084:	gap of	100 bp
*	134085	138474:	contig of 4390 bp in length	
*	138475	138574:	gap of	100 bp
*	138575	143774:	contig of 5200 bp in length	
*	143775	143874:	gap of	100 bp
*	143875	149669:	contig of 5795 bp in length	
*	149670	149769:	gap of	100 bp
*	149770	153536:	contig of 3767 bp in length	
*	153537	153636:	gap of	100 bp
*	153637	158023:	contig of 4387 bp in length	
*	158024	158123:	gap of	100 bp
*	158124	162445:	contig of 4322 bp in length	
*	162446	162545:	gap of	100 bp
*	162546	165243:	contig of 2698 bp in length	
*	165244	165343:	gap of	100 bp
*	165344	168875:	contig of 3532 bp in length	
*	168876	168975:	gap of	100 bp
*	168976	171973:	contig of 2998 bp in length	
*	171974	172073:	gap of	100 bp
*	172074	175989:	contig of 3916 bp in length	
*	175990	176089:	gap of	100 bp
*	176090	179781:	contig of 3692 bp in length	
*	179782	179881:	gap of	100 bp
*	179882	182243:	contig of 2382 bp in length	
*	182244	182343:	gap of	100 bp
*	182344	184565:	contig of 2222 bp in length	
*	184566	184665:	gap of	100 bp
*	184666	187147:	contig of 2482 bp in length	
*	187148	187247:	gap of	100 bp
*	187248	189548:	contig of 2301 bp in length	
*	189549	189648:	gap of	100 bp
*	189649	190661:	contig of 1013 bp in length	
*	190662	190761:	gap of	100 bp
*	190762	191967:	contig of 1206 bp in length	
*	191968	192067:	gap of	100 bp
*	192068	193487:	contig of 1420 bp in length.	
FEATURES	Location/Qualifiers			
source	1..193487			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="18"			
	/map="18p11.3"			
	/clone="Rp11-661013"			
misc_feature	1..15033			
	/note="assembly_fragment"			
misc_feature	15134...28850			
	/note="assembly_fragment"			
misc_feature	28951..43381			
	/note="assembly fragment"			

Thu Oct 25 13:09:01 2001

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 8, 2000 this sequence version replaced gi:10047674.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1313
Center clone name: 168_M_18
----- Summary Statistics

Sequencing vector: M13; M77815; 19% of reads
Sequencing vector: Plasmid; n/a; 81% of reads
Chemistry: Dye-terminator; 3% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 195941 bases at least Q40
Consensus quality: 202775 bases at least Q30
Consensus quality: 206217 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 208534; sum-of-contigs
Quality coverage: 24.4 in Q20 base.
NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 26180: contig of 26180 bp in length
26181 26280: gap of 100 bp
26281 27483: contig of 1203 bp in length
27484 27583: gap of 100 bp
27584 28811: contig of 1228 bp in length
28812 28911: gap of 100 bp
28912 30054: contig of 1143 bp in length
30055 30154: gap of 100 bp
30155 31369: contig of 1215 bp in length
31370 31469: gap of 100 bp
31470 32669: contig of 1200 bp in length
32670 32769: gap of 100 bp
32770 33804: contig of 1035 bp in length
33805 33904: gap of 100 bp
33905 35717: contig of 1813 bp in length
35718 35817: gap of 100 bp
35818 37338: contig of 1521 bp in length
37339 37438: gap of 100 bp
37439 39081: contig of 1643 bp in length
39082 39181: gap of 100 bp
39182 40202: contig of 1021 bp in length
40203 40302: gap of 100 bp
40303 41938: contig of 1636 bp in length
41939 42038: gap of 100 bp
42039 43493: contig of 1455 bp in length
43494 43593: gap of 100 bp
43594 44979: contig of 1386 bp in length
44980 45079: gap of 100 bp
45080 46806: contig of 1727 bp in length
46807 46906: gap of 100 bp
46907 48423: contig of 1517 bp in length
48424 48523: gap of 100 bp
48524 49581: contig of 1058 bp in length
49582 49681: gap of 100 bp
49682 51224: contig of 1543 bp in length
51225 51324: gap of 100 bp
51325 52909: contig of 1585 bp in length
52910 53009: gap of 100 bp

* 53010 54846: contig of 1837 bp in length
* 54847 54946: gap of 100 bp
* 54947 56941: contig of 1995 bp in length
* 56942 57041: gap of 100 bp
* 57042 60058: contig of 3017 bp in length
* 60059 60158: gap of 100 bp
* 60159 210734: contig of 150576 bp in length.
FEATURES
Location/Qualifiers
1. .210734
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="rp11-168M18"
/clone_lib="RPC1-11 Human Male BAC"
1. .26180
/note="assembly_fragment"
clone_end:SP6
vector_side:left
26281..27483
/note="assembly_fragment"
27584..28811
/note="assembly_fragment"
28912..30054
/note="assembly_fragment"
30155..31369
/note="assembly_fragment"
31470..32669
/note="assembly_fragment"
32770..33804
/note="assembly_fragment"
33905..35717
/note="assembly_fragment"
35818..37338
/note="assembly_fragment"
37439..39081
/note="assembly_fragment"
39182..40202
/note="assembly_fragment"
40303..41938
/note="assembly_fragment"
42039..43493
/note="assembly_fragment"
43594..44979
/note="assembly_fragment"
45080..46806
/note="assembly_fragment"
46907..48423
/note="assembly_fragment"
48524..49581
/note="assembly_fragment"
49682..51224
/note="assembly_fragment"
51325..52909
/note="assembly_fragment"
53010..54846
/note="assembly_fragment"
54947..56941
/note="assembly_fragment"
57042..60058
/note="assembly_fragment"
60159..210734
/note="assembly_fragment"
clone_end:m7
vector_side:right
BASE COUNT 57661 a 45068 c 43811 g 61994 t 2200 others
ORIGIN

Query Match 81.7%; Score 18.8; DB 63; Length 210734;
Best Local Similarity 90.9%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

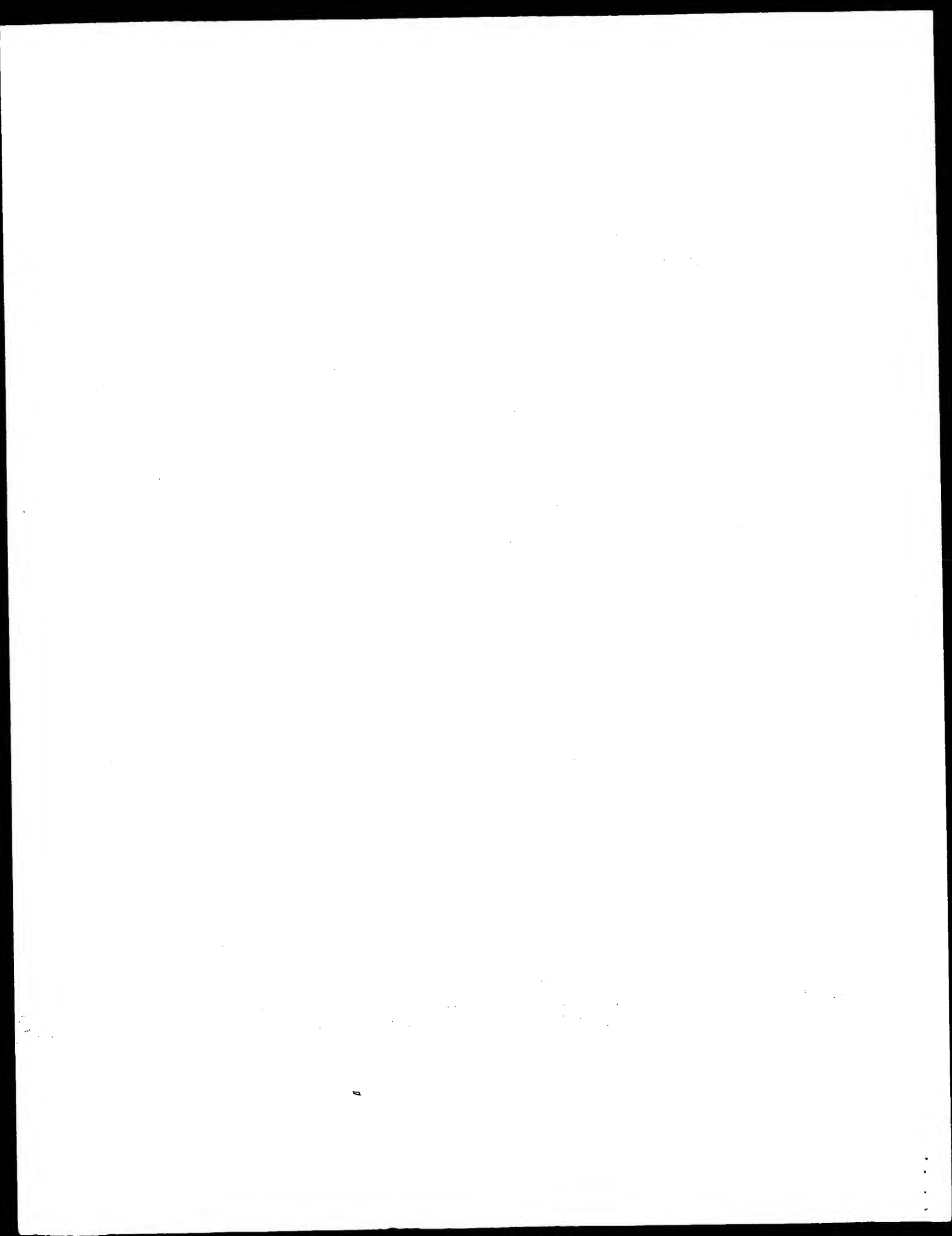
Thu Oct 25 13:09:01 2001

us-09-462-955-6.rge

Page 17

Qy 1 ctgcccaggccgaaggcctggg 22
||||| ||||| ||||| |||||
Db 203306 CTGCCAAGGCCCAAGGCCTGGG 203285

Search completed: October 24, 2001, 11:43:17
Job Time: 6382 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2001, 10:00:31 : Search time 180.6 Seconds
(without alignments)
24.109 Million cell updates/sec

Title: US-09-462-955-6

Perfect score: 23

Sequence: 1 ctgcccagggcgaagcctggga 23

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents.NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	Length	DB	ID	Description
1	18.2	79.1	3131	3	US-09-035-648-23		Sequence 23, Appl
2	18.2	79.1	3131	4	US-09-001-951-23		Sequence 23, Appl
c 3	17.8	77.4	2816	1	US-08-785-241-1		Sequence 1, Appl
c 4	16.6	72.2	1659	3	US-09-083-351-3		Sequence 3, Appl
c 5	16.6	72.2	1659	4	US-09-083-352-3		Sequence 3, Appl
c 6	16.6	72.2	2828	3	US-09-357-071-1		Sequence 1, Appl
c 7	16.6	72.2	3946	3	US-09-083-351-1		Sequence 1, Appl
c 8	16.6	72.2	3946	4	US-09-083-352-1		Sequence 1, Appl
c 9	16.4	71.3	4200	1	US-07-841-654B-1		Sequence 1, Appl
c 10	16.4	71.3	4200	1	US-07-946-234A-1		Sequence 1, Appl
c 11	16.4	71.3	4200	1	US-08-123-161A-1		Sequence 1, Appl
c 12	16.4	71.3	4200	1	US-08-483-278-1		Sequence 1, Appl
c 13	16.4	71.3	4200	5	PCT-US93-01560-1		Sequence 1, Appl
c 14	16.2	70.4	472	2	US-08-811-949-40		Sequence 40, Appl
c 15	16.2	70.4	747	1	US-08-257-341-6		Sequence 6, Appl
c 16	16.2	70.4	779	1	US-08-133-804-3		Sequence 3, Appl
c 17	16.2	70.4	779	1	US-08-461-838-3		Sequence 3, Appl
c 18	16.2	70.4	779	2	US-08-461-838-3		Sequence 3, Appl
c 19	16.2	70.4	942	3	US-08-732-412-1		Sequence 1, Appl
c 20	16.2	70.4	1065	1	US-08-427-640-1		Sequence 1, Appl
c 21	16.2	70.4	1065	1	US-08-427-640-5		Sequence 1, Appl
c 22	16.2	70.4	1065	2	US-08-811-949-60		Sequence 5, Appl
c 23	16.2	70.4	1068	1	US-08-137-116-2		Sequence 60, Appl
c 24	16.2	70.4	1068	1	US-08-427-640-3		Sequence 2, Appl
c 25	16.2	70.4	1068	2	US-08-811-949-44		Sequence 3, Appl
c 26	16.2	70.4	1068	2	US-08-811-949-46		Sequence 44, Appl
c 27	16.2	70.4	1068	2	US-08-811-949-52		Sequence 46, Appl
							Sequence 52, Appl

c 28	16.2	70.4	1068	2	US-08-811-949-58	Sequence 58, Appl
c 29	16.2	70.4	1068	6	5223256-3	Patent No. 5223256
c 30	16.2	70.4	1110	1	US-08-257-341-4	Sequence 4, Appl
c 31	16.2	70.4	1170	2	US-08-811-949-64	Sequence 64, Appl
c 32	16.2	70.4	1170	2	US-08-811-949-66	Sequence 66, Appl
c 33	16.2	70.4	1260	1	US-08-029-404-1	Sequence 1, Appl
c 34	16.2	70.4	1260	3	US-08-459-953A-1	Sequence 1, Appl
c 35	16.2	70.4	1314	2	US-08-811-949-50	Sequence 48, Appl
c 36	16.2	70.4	1314	2	US-08-811-949-50	Sequence 50, Appl
c 37	16.2	70.4	1314	2	US-08-811-949-54	Sequence 54, Appl
c 38	16.2	70.4	1314	2	US-08-811-949-56	Sequence 56, Appl
c 39	16.2	70.4	1419	2	US-08-811-949-62	Sequence 62, Appl
c 40	16.2	70.4	1461	1	US-08-403-634-1	Sequence 1, Appl
c 41	16.2	70.4	1461	4	US-08-913-441B-1	Sequence 1, Appl
c 42	16.2	70.4	1724	6	5200340-5	Patent No. 5200340
c 43	16.2	70.4	1738	6	5200340-1	Patent No. 5200340
c 44	16.2	70.4	1848	3	US-08-814-412-10	Sequence 10, Appl
c 45	16.2	70.4	1955	2	US-08-883-795A-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1

US-09-035-648-23
; Sequence 23, Application US/09035648
; Patent No. 6100031
; GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL

TITLE OF INVENTION: GROWTH AND PROLIFERATION

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/035,648

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/818,829

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Meiklejohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283

REFERENCE/DOCKET NUMBER: 07334/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 3131 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 449...2665

US-09-035-648-23

us-09-462-955-6.rni

Thu Oct 25 13:09:04 2001

Query Match 79.1%; Score 18.2; DB 3; Length 3131;
Best Local Similarity 87.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ctgccagccgagcagcctggga 23
||||||| ||||| ||||| ||
Db 957 CTGCCAGACCGATGGCTGAGA 979

RESULT 2
US-09-001-951-23
; Sequence 23, Application US/09001951
; Patent No. 6268470
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL
; TITLE OF INVENTION: GROWTH AND PROLIFERATION
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,951
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/818,829
; FILING DATE: 14-MAR-1997
; APPLICATION NUMBER: 60/013,438
; FILING DATE: 15-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3131 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 449...2665
; US-09-001-951-23

Query Match 79.1%; Score 18.2; DB 4; Length 3131;
Best Local Similarity 87.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ctgccagccgagcagcctggga 23
||||||| ||||| ||||| ||
Db 957 CTGCCAGACCGATGGCTGAGA 979

RESULT 3
US-08-785-241-1/C
; Sequence 1, Application US/08785241

Patent No. 5695963
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Russell, David W.
; APPLICANT: Tian, Hui
; TITLE OF INVENTION: Endothelial PAS Domain Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,241
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UTSD:1229
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2816 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-785-241-1

Query Match 77.4%; Score 17.8; DB 1; Length 2816;
Best Local Similarity 90.5%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctgccagccgagcagcctgg 21
||||||| ||||| ||||| ||
Db 2784 CTGCCAGGTAGAGGCTGG 2764

RESULT 4
US-09-083-351-3/c
; Sequence 3, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; TITLE OF INVENTION: TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HONG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible


```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1659 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-083-351-3

```

```

Query Match          72.2%; Score 16.6; DB 3; Length 1659;
Best Local Similarity 82.6%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 ctgcccaggccgaagcctggga 23
|||||  |||  |||  |||  |||
Db 743 CTGCCAGGGCGGGCGGGGA 721

```

```

RESULT 5
US-09-083-352-3/c
; Sequence 3, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1659 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-083-352-3

```

```

Query Match          72.2%; Score 16.6; DB 4; Length 1659;
Best Local Similarity 82.6%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 ctgcccaggccgaagcctggga 23
|||||  |||  |||  |||  |||
Db 743 CTGCCAGGGCGGGCGGGGA 721

```

```

RESULT 6
US-09-357-071-1
; Sequence 1, Application US/09357071
; Patent No. 6043091
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF LIVER GLYCOGEN PHOSPHORYLASE EXPRESSION
; FILE REFERENCE: RTS-0074
; CURRENT APPLICATION NUMBER: US/09/357,071
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 2828
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (114)..(2657)
US-09-357-071-1

```

```

Query Match          72.2%; Score 16.6; DB 3; Length 2828;
Best Local Similarity 82.6%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 1 ctgcccaggccgaagcctggga 23
|||||  |||  |||  |||  |||
Db 1191 ctgcccctgtccaagcgtgga 1213

```

```

RESULT 7
US-09-083-351-1/c
; Sequence 1, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351

```

FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 475..2133
US-09-083-351-1

Query Match 72.2%; Score 16.6; DB 3; Length 3946;
Best Local Similarity 82.6%; Pred. No. 87;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctgcccagggccgaagcctggga 23
||||| ||| ||| |||
DB 1217 CTGCCAGGGCGCGCGGGGA 1195

RESULT 8
US-09-083-352-1/c
Sequence 1, Application US/09083352
Patent No. 6207450
GENERAL INFORMATION:
APPLICANT: Sheffield, Val C.
APPLICANT: Alward, Wallace L.M.
APPLICANT: Stone, Edwin M.
APPLICANT: Nishimura, Darryl
APPLICANT: Patil, Shiva
TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,352
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3946 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 475..2133
US-09-083-352-1

Query Match 72.2%; Score 16.6; DB 4; Length 3946;
Best Local Similarity 82.6%; Pred. No. 87;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctgcccagggccgaagcctggga 23
||||| ||| ||| |||
DB 1217 CTGCCAGGGCGCGCGGGGA 1195

RESULT 9
US-07-841-654B-1/c
Sequence 1, Application US/07841654B
Patent No. 5260209
GENERAL INFORMATION:
APPLICANT: Campbell, Kevin P.
APPLICANT: Idraghimov-Beskrovnya, Oxana
APPLICANT: Ervasti, James M.
APPLICANT: Leveille, Cynthia J.
APPLICANT: Matsumura, Kiichiro
TITLE OF INVENTION: DNA ENCODING DYSTROPHIN-ASSOCIATED
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/841,654B
FILING DATE: 19920220
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: UIR89-11AA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 861-6240
TELEFAX: 617 861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4200 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 170..2855
US-07-841-654B-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0;

QY 6 caggccgaagcctggga 23
||||| ||| ||| |||
DB 4051 CAGGCCCAAGCGCTGGGA 4034

```
RESULT 10
US-07-946-234A-1/c
; Sequence 1, Application US/07946234A
; Patent No. 5308752
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Matsumura, Kiichiro
; TITLE OF INVENTION: DIAGNOSIS OF AUTOSOMAL MUSCULAR DYSTROPHY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/946,234A
; FILING DATE: 19920914
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: UIRF89-11AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)861-6240
; TELEFAX: (617)861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
US-07-946-234A-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 caggccgaaggcctggga 23
||||| |||||||
Db 4051 CAGGCCCAAGGCTGGGA 4034

RESULT 11
US-08-123-161A-1/c
; Sequence 1, Application US/08123161A
; Patent No. 5449616
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Roberts, Steven L.
; APPLICANT: Anderson, Richard D.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
```

```
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,161A
; FILING DATE: 16-SEP-93
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/946,234
; FILING DATE: 14-SEP-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF89-11A4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..2855
US-08-123-161A-1

Query Match 71.3%; Score 16.4; DB 1; Length 4200;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 caggccgaaggcctggga 23
||||| |||||||
Db 4051 CAGGCCCAAGGCTGGGA 4034

RESULT 12
US-08-483-278-1/c
; Sequence 1, Application US/08483278
; Patent No. 5686073
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Ervasti, James M.
; APPLICANT: Leveille, Cynthia J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,278
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,161
```

```

; LOCATION: 170..2855
PCT-US93-01560-1

Query Match 71.3%; Score 16.4; DB 5; Length 4200;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 caggccgaaggcctggga 23
||||| |||||||||
Db 4051 CAGGCCCAAGGCTGGGA 4034

RESULT 14
US-08-811-949-40/c
; Sequence 40, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..471
; US-08-811-949-40

Query Match 70.4%; Score 16.2; DB 2; Length 472;
Best Local Similarity 85.7%; Pred. No. 1.2e-02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgcccaggccgaaggcctggg 22
||||| |||||||
Db 70 TGCCCAAGCCAGTGCTGGG 50

RESULT 15

```

US-08-257-341-6
 ; Sequence 6, Application US/08257341
 ; Patent No. 5525491
 ; GENERAL INFORMATION:
 ; APPLICANT: HUSTON, JAMES S
 ; APPLICANT: OPPERMAN, HERMANN
 ; APPLICANT: TIMASHEFF, SERGE N
 ; TITLE OF INVENTION: SERINE RICH PEPTIDE LINKER
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CREATIVE BIOMOLECULES, INC./PATENT DEPT.
 ; STREET: 35 SOUTH STREET
 ; CITY: HOPKINTON
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 01748
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/257,341
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/842,149
 ; FILING DATE:
 ; APPLICATION NUMBER: US 07/662,226
 ; FILING DATE: 27-FEB-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CAMPBELL ESQ, PAULA A
 ; REGISTRATION NUMBER: 32,503
 ; REFERENCE/DOCKET NUMBER: CRP-064CP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/248-7000 (ATTY)
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 747 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: syn DNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..747
 ; US-08-257-341-6

Query Match 70.4%; Score 16.2; DB 1; Length 747;
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tgcccagccgaagcctggg 22
 ||| |
 Db 645 TGTCGAGCCGAGAGACTGGG 665

Search completed: October 24, 2001, 10:00:32
 Job time: 217 sec

